Alberta Science Achievement Study

APRIL, 1979

Alberta

Minister's Advisory Committee
on Student Achievement

ALBERTA SCIENCE ACHIEVEMENT STUDY

A STUDY CONDUCTED FOR
THE MINISTER'S ADVISORY COMMITTEE ON
STUDENT ACHIEVEMENT

Condensed Report

- by -

Morris Treasure
Planning and Research Branch
Alberta Education

Project Completion Date:
Fall, 1978

Table of Contents

Table of Contents (cont.)

List of Tables

PART I

List of Tables

PART II

List of Tables

PART II (cont.)

List of Tables

PART II (cont.)

List of Tables

PART II (cont.)

ABSTRACT

In July 1977 Dr. M. R. Treasure was commissioned by the Minister's Advisory Committee on Student Achievement to study the levels of student achievement in the field of science at grades three, six, nine and twelve. The investigator was to provide baseline information for future assessments and a summary of student achievement of curricular objectives.

The grades three and six tests were constructed from items culled from existing item banks with established characteristics, such as those of the National Assessment of Educational Progress. The tests were administered using an item-examinee sampling scheme to elementary students in a 10 percent stratified sample of schools across Alberta on May 17, 1978.

The grades nine and twelve tests were an amalgamation of the STEP II Science Test, the Test of Understanding Science (TOUS), form Jw, and supplementary items related to specific course objectives not tested by STEP II or TOUS. The tests were administered to secondary students in a 10 percent stratified sample of junior and senior high schools across Alberta on May 17, 1978.

Between 2000 and 3000 students were tested at each grade level. Student performances on each cluster of items, which appear in this report, were generally satisfactory. But some specific areas of weakness were identified: knowledge of scientific methods in grade three; knowledge of physical science in grades three, six and twelve; and earth-space science and general knowledge of science and scientists in grades nine and twelve. The grades nine and twelve performances on the standardized STEP II Science for 2A and 3A were better than the U.S. norms established for those grades.

To speculate without facts is to attempt to enter a house of which one has not the key by wandering aimlessly round and round, searching the walls, now and then peeping through the windows. Facts are the key.

- Julian Huxley
Essays in Popular Science

Preface

The Minister's Advisory Committee on Student Achievement (MACOSA) was established by ministerial order in October 1976 in response to growing concerns expressed by the public at large, government, labor, business, students and educators regarding the quality and standards of basic education in Alberta.

MACOSA commissioned a number of studies, primarily to provide basic information for a summary of current levels of achievement in Alberta and to provide baseline data for future assessment. These studies fell into three categories: (1) preliminary studies, (2) achievement studies, and (3) other studies.

This achievement study, the *Alberta Science Achievement Study*, was designed to provide information about current levels of achievement in science among students in Alberta schools and to provide a data base for future assessments.

This report, which represents the findings and conclusions of the researcher, was presented to MACOSA as information.

ACKNOWLEDGEMENTS

The advice and assistance of a wide group of colleagues in Alberta Education and the university community was a valuable prerequisite of the study.

The assistance of Dr. J. E. Reid and the staff of the Student Evaluation and Data Processing Branch is publicly acknowledged. The support and assistance of the Planning and Research Branch is also greatly appreciated.

Advice, sympathy and assistance far beyond the usual responsibilities of a departmental committee were received from the following members of the steering committee:

Dr. L. R. Tolman, Chairman
Dr. H. C. Rhodes, Recorder
Dr. W. G. Holliday
Mr. M. Lynch
Mr. I. Ibuki
Dr. H. G. Sherk, Executive Secretary of MACOSA

The cooperation of the participants in the study is also greatly appreciated--the students who took the time to write the tests, the teachers who administered the tests, and those who cared enough to respond in writing. The investigator also acknowledges the assistance of the support staff of the Planning and Research Branch who have typed thousands of pages of tests, charts and reports.

PART I

INTRODUCTION

1.1 Background to the Study

At its meeting of July 13, 1977, the Minister's Advisory Committee on Student Achievement (MACOSA) approved the proposal to assess student achievement in the field of science at grades three, six, nine and twelve. The study was to be completed by September 30, 1978.

The purpose of the study was to investigate the levels of student achievement in science in Alberta. The assessment was to be in terms of the educational goals as outlined in the program of studies statements for Alberta schools. The data gathered during the study was to provide base-line information for future assessments. A true assessment of the variation of student achievement over time can only be made after a repetition of the assessment.

To direct the progress of the study and to provide advice to the investigator on the conduct of the study, a steering committee of five members was established by MACOSA on May 11, 1977. This committee's membership consisted of Dr. L. Tolman, Mr. I. Ibuki, Mr. M. Lynch, and Dr. W. Holliday. Dr. H. C. Rhodes served as recorder. The first meeting was held on June 27, 1977.

1.2 Objectives of the Study

The study was to provide a clear and concise picture of the current level of the science competencies of Alberta students.

The tests used to collect the data were to have content validity; that is, they were to reflect the Alberta science curriculum at each level.

1.3 The Problem

The problem assigned by MACOSA was to assess current levels of general scientific knowledge in grades three, six, and nine as defined by the provincial curriculum guidelines. The assessment at grade twelve was to be a more general survey which would not measure the content of any specific science course. The assessment program was to provide a descriptive data base about current levels of scientific knowledge among Alberta students which could be used as a base-line for future assessments.

1.4 Experience Elsewhere

The two principal sources of information about the widespread testing of students were the National Assessment of Educational Progress (N.A.E.P.) from the United States and the British Columbia Learning Assessment Program. The reports from these two assessments in particular provided much encouragement and gave the researcher confidence that the task was manageable.

1.5 Approach to the Problem

The probelm was attacked by dividing it into its constituents; that is, the bounds of the assessment endeavour were determined at each of the four grade levels. Five phases of the study were identified for each of the four grade levels to be tested: curricular search, item or test identification, test construction, test administration, and results interpretation.

1.5.1 Extent of and Constraints Upon the Study

The steering committee was directed to limit the study to a 10 percent sample of students in grades three, six, nine and twelve. The assessments for grades three, six and nine were to be curriculum valid and the assessment at grade twelve was to focus more on the general goals for secondary science education.

Some of the constraints upon the study were planned; others appeared as a consequence of the decisions taken while the study was being conducted. At each grade level the study focussed on a selected number of objectives which were from the immediately preceding grades as well as from the target grade and were representative of the wide range of objectives taught during the course of a year.

The study relied on pencil and paper tests for the student data, with all of the advantages and disadvantages that this mode implies. The problem of reading difficulty at the grade three level was controlled by having teachers read the questions and responses to the students.

The study was to be designed, conducted and reported in a fourteen-month period from July 15, 1977 to September 30, 1978. Choices, such as the use of an item bank tailored to the Alberta science curriculum, were severely restricted.

Time constraints proved to be decisive in limiting the steering committee's choice of testing modes.

1.5.2 Curricular Bases For the Assessment

Program of studies statements for each of the four levels were carefully examined. The grades three and six programs of studies were based on the six conceptual schemes first outlined in the N.S.T.A. publication, Theory into Action (1964). These program statements of the science curriculum have been in place since 1968, but revision of the elementary science program is under way. The steering committee for the present study advised that the current Program of Studies for Elementary Schools--1975 should provide the organizing framework for the content of the test.

The junior high school program of studies statement (1975) was closely examined as well, and was deemed to be acceptable as an organizing construct for a test of student achievement.

The assessment at the grade twelve level was to be less oriented to specific program content and more related to global understandings of science and the scientific enterprise in general.

At each of the four divisions of the program for grades one through twelve, the program of studies statement formed the organizing construct for student assessment. The specific program content bounded the grades three, six and nine tests and the secondary school science objectives defined the grade twelve science test.

1.5.3 Items Used in the Assessment

The steering committee for the science study determined that, given the time constraints and the present state of the program, the procedure for the grades three and six assessment should be based on a bank of items.

The initial target was to obtain or develop 120 items for each of the grades three and six assessments. These items were to be representative of the six conceptual schemes and cover the broad areas of knowledge of science, attitudes toward science as a process, science as a field of human endeavour, and attitudes toward scientists as individuals.

After due consideration of the parameters surrounding the study, the steering committee accepted that a commercial achievement test could be used as part of the assessment at grades nine and twelve. This decision was made on two main grounds: one, that the shortage of time precluded developing an item bank and adequately piloting the instruments for grades nine and twelve; and two, that the content of the commercial test would be somewhat consistent with the curriculum in junior high, since the Alberta program was very similar to that in use elsewhere in North America. It was also observed

that the purpose of the grade twelve assessment (a general reading of science achievements) could be adequately served by using an appropriate commercial test. In each case, additional items were developed to improve the test's match with the curriculum.

1.5.4 The Tests

Tables of specifications for the grades three and six tests were developed to guide construction of the tests. After piloting the tests, the test developer dropped a number of items. The final structure of the tests is shown in Tables 1-1, 1-2, 1-3 and 1-4.

Table 1-1

GRADE 3 ITEMS CLASSIFIED BY PROGRAM OBJECTIVES

Specific Science Content	Knowledge	Comprehension	Application	Higher
1. Conservation of Energy	3	5	3	-
2. Conservation of Matter	3	4	3	-
3. Interdependence of Living Things	2	2	2	3
4. Living Things--Heredity and Environment	2	7	3	-
5. Living Things and Change	2	3	-	-
6. Earth-Space Science	3	3	-	-
7. Scientists as People and Researchers	9			
8. Methods of Science - Observing	11			
- Classifying	7			
- Manipulation of Data	7			
- Experimenting	1			
- Measuring	8			
- Inferring	7			
9. The Scientific Enterprise	5			
10. Interest in Science	6			

Table 1-2

GRADE 6 ITEMS CLASSIFIED BY PROGRAM OBJECTIVES

<table>
<tr><th>Specific Science Content</th><th>Knowledge</th><th>Compre-hension</th><th>Applica-tion</th><th>Higher</th></tr>
<tr><td>1. Conservation of Energy</td><td>7</td><td>10</td><td>7</td><td>-</td></tr>
<tr><td>2. Conservation of Matter</td><td>5</td><td>6</td><td>3</td><td>-</td></tr>
<tr><td>3. Interdependence of Living Things</td><td>8</td><td>5</td><td>6</td><td>1</td></tr>
<tr><td>4. Living Things--Heredity and Environment</td><td>6</td><td>4</td><td>5</td><td>-</td></tr>
<tr><td>5. Living Things and Change</td><td>4</td><td>4</td><td>-</td><td>2</td></tr>
<tr><td>6. Earth-Space Science</td><td>4</td><td>4</td><td>3</td><td>-</td></tr>
<tr><td>7. Scientists as People and Researchers</td><td colspan="4">7</td></tr>
<tr><td>8. Methods of Science - Observing</td><td colspan="4">1</td></tr>
<tr><td>- Manipulation of Data</td><td colspan="4">5</td></tr>
<tr><td>- Experimenting</td><td colspan="4">3</td></tr>
<tr><td>- Measuring</td><td colspan="4">9</td></tr>
<tr><td>- Inferring</td><td colspan="4">6</td></tr>
<tr><td>9. Scientific Enterprise</td><td colspan="4">7</td></tr>
<tr><td>10. Interest in Science</td><td colspan="4">12</td></tr>
</table>

The item bank for grades three and six was developed from two main sources: the N.A.E.P. Science Technical Report (1977) and the British Columbia Assessment Study. Some 250 items were accumulated and assigned to one of the ten clusters in either grade three or grade six. No minimum number of items were targeted at either grade level but an attempt was made to provide a surplus from which items could be culled.

On the basis of an initial pilot, in which a minimum of 100 students tried each item, a number of items were dropped from further consideration because of technical inadequacy.

The 114 grade three items that were left after piloting were randomly assigned to four tests, with one exception--items related to a specific diagram were assigned to the same test. To reduce the impact of the testing program upon instructional time, the steering committee established a goal of limiting the time required from each pupil to about 20 to 30 minutes.

The 144 grade six items which remained after the piloting procedure were also randomly assigned to four tests (A, B, C and D). Care was taken to group items which were related.

To provide a small measure of change in student response, particularly in attitude toward science and stereotypes of scientists, 47 of the items were used on both tests.

In designing the test for the secondary science program, the researcher recognized that the tests could only sample concepts from the various subject areas and thought levels. The plan of the test is based upon the curricula of the grades to be tested by the instruments. Tables 1-3 and 1-4 illustrate the distribution of the items on the tests. It should be restated that the blueprints were developed before the various commercial tests were examined and the choices were made on the basis of a "best fit". The voids in the blueprints were then filled by additional items from various sources. The charts show the number of items that test each grade level's concepts. In addition, the items that test across grade levels are shown as are the 45 items that form the TOUS test. The items are also categorized according to the mental reasoning required in their solution.

Table 1-3

GRADE 9 ITEMS CLASSIFIED BY
PROGRAM OBJECTIVES

Specific Science Content	Knowledge	Compre-hension	Applica-tion	Higher
1. Life Science Concepts	4	1	5	2
2. Earth-Space Concepts	5	3	8	1
3. Physical Science Concepts	4	3	8	2
4. Methods of Science	-	9	8	2
5. Scientific Enterprise	8			
6. Scientists	11			
7. Practice of Science	26			

The grade nine test was made up of the Sequential Test of Educational Progress, Series II (STEP II), form 3A; 15 additional items related to specific curricular concepts not included in STEP II; and the Test of Understanding Science (TOUS), form Jw. This resulted in a 110-item test that was first piloted with 213 students in eight classrooms in Edmonton and a number of rural points. The pilot results indicated problems with three items; these were modified and the test was prepared for administration.

Table 1-4

GRADE 12 ITEMS CLASSIFIED BY PROGRAM OBJECTIVES

Specific Science Content	Knowledge	Comprehension	Application	Higher
1. Biology Concepts	6	4	6	-
2. Physics Concepts	3	8	3	1
3. Chemistry Concepts	6	3	8	-
4. Earth-Space Concepts	1	-	2	-
5. Methods of Science	1	15	13	7
6. Scientific Enterprise		8		
7. Scientists		11		
8. Practice of Science		26		

The grade twelve test was similarly compiled from the STEP II, form 2A; 12 additional items; and the 45-item TOUS, form Jw. The resulting 132-item test was piloted with 188 students in test schools across Alberta. Only a few minor problems appeared; the items were then revised and the test was prepared for administration.

There was less concern at the grades nine and twelve levels about the time required to write the test, but an attempt was made to limit the time to one 80-minute period.

1.5.5 Validation

The initial choice of items at the grades three and six levels was done by the study researcher. These items were randomly assigned to four tests, which were then examined by teachers at each grade level in ten schools. The pilot schools were chosen at random from a list of schools that would not be involved in administering the MACOSA tests on May 17.

Responses were received from 20 primary teachers, two from each of the randomly selected schools. Most of the comments were related to specific items, and many comments resulted in alteration or, in some cases, deletion of the item. A few

comments were related to procedures, such as the use of IBM answer sheets and the problem of reading difficulty. These comments resulted in having teachers read the items to grade three students.

Eighteen responses were received from grade six teachers, two from each of the randomly selected schools. Most of the responses resulted in modifications of items, particularly in simplifying diagrams and changing the wording of items.

The choice of STEP was made on the basis of reviews of the test in the Fifth Mental Measurements Yearbook, the Report of the Tests for Alberta Schools, and the judgements of the steering committee.

The STEP II, form 2A and 3A tests were ordered from the publisher, the authors of the TOUS form Jw were contacted and permission was granted to print in the required quantities, and the additional items requested by the steering committee were developed. The battery of tests were then distributed to ten junior high schools and ten senior high schools across the province for administration and teacher comments.

Five responses were received from grade nine teachers and two from grade twelve teachers. These comments were mainly related to the use of the TOUS test as a measure of knowledge about the scientific enterprise. The comments were answered, not so much by changing the tests as by including an explanatory note in the information package accompanying the tests.

The choice of the Test On Understanding Science (TOUS) was based on its use in a number of studies in Alberta. The TOUS was developed at the Harvard University Graduate School of Education. It was developed to measure the intangible aspects of science; that is, knowledge about science and scientists.

The TOUS was developed around themes in three areas: understandings about the scientific enterprise; scientists; and the methods and aims of science. The test items were subjected to intensive scrutiny and testing with panels of science educators, teachers and scientists. They were then validated by administration to 3000 high school students. Tentative norms have been established. A test-retest study was conducted with a group of 78 talented high school students working with active scientists in two summer programs. The students' test scores improved significantly at the end of the summer program, whereas a similar group of high school students not participating in the summer program failed to register significant gains. The test-retest reliabilities have not been reported. Until a more complete analysis and report of the wider use of TOUS has been made, the test as a whole should be used cautiously in applying the test results to individuals.

1.5.6 Matrix-Sampling (Grades Three and Six)

To reduce the size of the task being demanded of an individual student and to reduce the magnitude of the total testing program at a given grade level, a scheme of item sampling was combined with a stratified school sampling technique.

The 114 items in the grade three item bank were randomly assigned to one of four test booklets. Each of the schools selected to participate in the study was then sent as many booklets as it had students enrolled in grade three. Similarly, the 144 items in the grade six item bank were randomly assigned to one of four test booklets and each participating school was sent as many booklets as it had students enrolled in grade six.

In grades nine and twelve the steering committee felt that every student in the study could complete the entire test and that there was little need for the matrix sampling technique.

1.5.7 Student Sample

The sample of students was selected by the Student Evaluation and Data Processing Branch of Alberta Education. A 10 percent stratified sample of schools was selected from the current list of operating schools. The grades three and six stratification was two-way, by type of school system and by size of enrolment in grades three and six. The types of school systems included large urban, small urban, and rural (counties, divisions and independent school districts). The enrolment categories varied slightly for elementary and secondary but they were generally small to large. Grades three and six school enrolment categories were: <20, 21-30, 31-60 and >60. The grades nine and twelve samples were stratified only by class size since the number of schools in each sub-group began to be very small. The grade nine categonies were: <20, 21-30, 31-60, 61-100, 101-150 and >150. The grade twelve categories were: <30, 31-60, 61-100 and >100.

The school sample consisted of 101 schools offering grade three; 96 schools offering grade six; 40 schools offering grade nine; and 24 schools offering grade twelve.

1.6 Structure of the Study

A preliminary study of the problem was commissioned by MACOSA at its third meeting in February, 1977. The researcher was directed to prepare detailed plans for the study at the fifth meeting in May, 1977. These were accepted at the sixth meeting in July, 1977 and a steering committee was subsequently established. The study was to be completed and a draft report submitted by September 30, 1978.

1.6.1 Item Specification and Collection

The steering committee directed that a curriculum-valid item bank of some 240 items be developed for the test for grades three and six. The researcher was also directed to review the available commercial tests to find ones that could be used to adequately test student performance in grades nine and twelve.

The items for the elementary grades were chosen from those used in student assessments by the National Assessment of Educational Progress in the United States, the International Association for the Evaluation of Educational Achievement for UNESCO, and the British Columbia Student Assessment Program.

1.6.2 Communications

A letter from the chairman of MACOSA, Dr. J. S. T. Hrabi, was sent to all participating school principals and their superintendents early in February asking that the May 17 date be reserved for testing purposes. This was followed with a letter to the principals and participating teachers on April 13 advising them specifically of the test and grade level involved. This letter asked that the numbers of students in the participating classes be reported to the researcher as soon as possible.

1.6.3 Data Analysis

The answer sheets were returned to Alberta Education and scored on the optical scorer by mid-June. The processing could not, however, be completed until mid-July. The item data was first processed by the department's item analysis program and then by the item sort program.

1.7 Test Results

Table 1-5

BASE-LINE RESULTS - INITIAL ADMINISTRATION

Grade	N	# Of Items	Average Raw Score	Average %	Standard Deviation (Raw Score)
12	2125	132	78.9	59.8	18.2
9	2426	110	62.2	56.6	13.0
6	2935	144	90.9*	63.6	9.4**
3	3073	114	68.3*	59.9	7.3**

* Pooled mean raw score
** Pooled estimate of the standard deviation

The test statistics are difficult to interpret, in that both the grades nine and twelve tests are composed of two quite different tests. The grades three and six tests also show the effect of mixing knowledge-based items with opinion items. For truly informative data one should look at the sub-test and item cluster summaries. An important product of these assessment procedures is that it is now possible to compare future administrations of the STEP II, forms 2A and 3A, with Alberta norms.

1.7.1 Elementary Science Tests

At each of the tested grade levels the test items were grouped or clustered according to the content of the question. The grades three and six content clusters were based on the six major conceptual schemes which define the elementary program.

Drawing inferences from different rankings in equivalent clusters on the grades three and six tests must be done cautiously and with reservation because different test items measuring the same objectives cannot be equated with any precision.

Table 1-6

GRADE 3 OBJECTIVE-BASED ITEM CLUSTERS

Cluster	# Of Items	Average Student Scores (%)	Rank
1. Physical Science	11	53.2	9
	10	57.3	8
2. Life Science	9	63.6	5
	12	66.0	3
	5	58.2	6
3. Earth-Space Science	6	77.2	1
4. Scientists	9	65.5	4
5. Science Methods	41	57.4	7
6. Aims of Science	5	51.7	10
7. Interest In Science	6	66.1	2
	114		

Table 1-7

GRADE 6 OBJECTIVE-BASED ITEM CLUSTERS

Cluster	# Of Items	Average Student Scores (%)	Rank
1. Physical Science	24	60.0	8
	14	58.8	9
2. Life Science	20	66.5	4
	15	65.0	6
	10	47.8	10
3. Earth-Space Science	11	70.8	2
4. Scientists	7	73.0	1
5. Science Methods	24	65.8	5
6. Aims of Science	7	69.2	3
7. Interest In Science	12	61.2	7
	144		

The first theme is a study of the physical science concept of the conservation of energy. The average grade six student scores were relatively the same as those for students in grade three. This item cluster ranked ninth out of ten for the grade three student and ninth out of ten for grade six.

The second theme is related to the first and is a study of the conservation of mass. The performance of both groups of students on items related to this concept was quite similar, with the grade three performance ranking eighth out of ten. Both of these physical science conceptual themes include very abstract ideas that are not readily observable by students.

The next three item clusters deal with life science topics. The third conceptual scheme deals with the interdependence among living things and their environment. Student performance again proved to be similar at both grade levels, with the grade six student performance ranking this item cluster fourth out of ten, and the grade three students ranking this cluster fifth out of ten.

The fourth theme deals with the organism as a product of heredity and environment. The student performances had similar raw scores. The grade three student performance ranked this cluster third, and the grade six student performance ranked it fifth out of ten.

The earth-space content of theme six seemed to attract the attention of students at both levels. Student performance ranked this theme first at grade three and second at grade six. Grade six students performed better on the common items related to this conceptual scheme, as one might predict.

The seventh item cluster contains those items which call for a knowledge and understanding of scientists. Grade six students tend to have more of the view shared by adults than do grade three students. On the seven common items the grade six performance was significantly better than that of the grade three students.

The content of the items in the cluster called "knowledge and understanding about the methods and processes of science" was not understood as well at the grade three level as at the grade six level (seventh out of ten as opposed to sixth out of ten). The process dimension is not as evident as it is later on in the program.

The ninth item cluster, items that deal with knowledge and understanding of the scientific enterprise, was answered far better by grade six students than by grade three students. Perhaps exposure to a wider range of experiences would account for the improvement in student performance.

1.7.2 Common Item Results

Forty-seven items were common to both the grades three and six tests--18 items from each of the content areas of grades one to three, and the same number from grades four to six. A further 11 items measured interest in science and opinions and beliefs about scientists. The relative student performance is shown in Table 1-8.

Table 1-8

AVERAGE STUDENT PERFORMANCE
ON THE COMMON ITEMS

	Student Performance (%) Grade 3	Grade 6
Grade 3 Target Items (18)	63.0	77.0
Grade 6 Target Items (18)	49.4	59.1
Interest And Opinion Items (11)	61.4	74.0
TOTAL	57.4	69.4

The differences in student performance is as expected, with a somewhat low performance by the grade six students on the grade six items. In general, the grade six student performance was substantially higher than that of the grade three students.

In the content areas the performance pattern is not as consistent nor as clear. Table 1-9 shows the relative performance on the common items clustered by item content.

Table 1-9

AVERAGE STUDENT PERFORMANCE ON COMMON ITEMS
IN THE CONTENT CLUSTERS

Content	# Of Common Items	Student Performance (%) Grade 3	Grade 6
Physical Science	5	62.9	64.3
Life Science	10	58.2	74.2
Earth Science	4	73.1	86.9
Processes & Methods	13	48.4	58.1
Scientists	10	57.5	74.9
Interest in Science	5	61.1	70.0

1.7.3 Secondary Science Tests

Table 1-10

SUMMARY OF STEP II TEST RESULTS
(RAW SCORES)

	Form 3A (Grade 9)		Form 2A (Grade 12)	
	Alberta (Spring '78)	U.S.A. (Spring '70)	Alberta (Spring '78)	U.S.A. (Spring '70)
N	2426	2637	2125	2285
Average Score	34	32	47	42
Standard Deviation	7	12	12	13

The most substantial differences between the average grade three and grade six student performances are on those item clusters dealing with earth science, life science and opinions and beliefs about scientists. The smallest differences are in those item clusters dealing with physical science, methods of science, and interest in science.

The junior high and senior high school science tests were composed of the STEP II, form 3A and form 2A; additional items related to specific content areas not covered by STEP; and the TOUS, form Jw.

The average grade twelve student response on STEP II, 2A, was 47 out of 75 items, as compared with the United States norming population results of 42 out of 75.

The average student response on STEP II, 3A, was 34 out of 50 items, as compared with the United States norming results of 32 out of 50 for grade nine. The comparisons are to be evaluated cautiously, because the American norming population was structured to approximate United States school population characteristics. However, accepting that the populations are different, the Alberta students out-performed their U.S. counterparts ($p < 0.05$).

Student performance on the STEP followed the generally accepted belief that boys achieve better than girls in science. However, the literature does not offer a convincing explanation for this difference. The difference of 1.93 raw score points is statistically significant for the numbers

involved. The average student performance in schools of different sizes, however, showed no statistical significance, and the question remains unresolved. The question of the effect of school size upon student achievement in science would have to be investigated further with other variables included such as the amount of class time devoted to science, the availability and use of laboratory space, scholastic aptitude of the students, and so on.

Table 1-11

SUMMARY OF TOUS, FORM Jw RESULTS
(RAW SCORES)

	Grade 9			Grade 12		
	Boys	Girls	Total	Boys	Girls	Total
N	1234	1144	2426	1059	1034	2125
Average Score	21.1	22.6	21.8	24.6	25.8	25.3
Standard Deviation	6.1	5.8	6.0	6.7	6.2	6.5

The differences between boys and girls are significant ($p < 0.001$).

Form Jw of TOUS has 45 items grouped in three clusters: understandings about the scientific enterprise, understandings about scientists, and methods of science. The results from this test must be used with caution since it is primarily a research instrument. It is, however, one of the few tests in this general area of non-subject-specific science testing. It is interesting to note that girls did significantly better ($p < 0.05$), than did boys on TOUS at both the grades nine and twelve levels.

Some concern has been expressed by various teachers on the reviewing panels for the MACOSA testing program that the non-cognitive items were an inappropriate part of the science achievement testing program. These comments can only be received and reported. However, the scope and sequencing of the tests by the steering committee has been in terms of the curriculum for the various grades as made explicit in the program of studies statements. These statements contain both cognitive and non-cognitive objectives.

The response pattern of grade nine boys and girls on the TOUS was the same as in grade twelve, except that the achieve-

ment levels of the grade twelve students were understandably higher. A survey of the research literature does not lead one to a cogent argument for the observed difference in the achievement levels of boys and girls.

Table 1-12

GRADE 9 OBJECTIVE-BASED ITEM CLUSTERS

Cluster	Rank	Average (%)	# Item
Life Science	1	69.3	12
Earth Science	4	56.7	17
Physical Science	2	62.6	17
Methods of Science (Processes)	3	62.4	19
Science as a Human Endeavour	5	49.2	45

Table 1-13

GRADE 12 OBJECTIVE-BASED ITEM CLUSTERS

Cluster	Rank	Average (%)	# Item
Biology	1	72.7	16
Earth Science	6	47.9	3
Physics	4	56.9	14
Chemistry	5	55.1	17
Methods of Science (Processes)	2	64.8	36
Science as a Human Endeavour	3	57.3	45

An examination of the item clusters for secondary science tests revealed that there were some similarities in the pat-

tern of student responses to essentially parallel tests for grades nine and twelve. Both groups performed best on the life science items and below average on the earth science items.

The process dimension of the program appears to show up fairly well. The student performance in grade nine ranked this cluster of items third of seven. The high school student performance ranked this cluster second of eight.

In junior high, for the items that dealt with the grade nine content, physical science showed up second of seven in terms of student performance. In the senior high school student performance rankings, the biology content questions were handled better by the general population of students, with physics a distant second and chemistry third.

In examining the grade twelve results, it should be kept in mind that the items are not related to specific courses but rather attempt to measure a general knowledge of various scientific concepts, procedures and opinions. The subject area breakdown is given here only to provide a point of comparison with the more subject-specific grade nine results.

As in any attempt to measure student achievement from a focus far removed from the classroom, one should remember the differences resulting from the translation of program of studies statements to classroom behaviors of teachers and from the independent translation of the assessment team. These translations may result in a close match between learning and testing or they may result in quite differing dimensions being examined. To attempt to keep the match as close as possible, great care was taken in item selection, piloting and analysis.

1.8 Findings

The findings of the science study at the grade three level were based on 114 items chosen from a variety of sources and validated with a randomly selected group of 20 primary teachers. At the grade six level the findings are based on 146 items chosen sources similar to those of the grade three items. The items were validated by a panel of 18 upper elementary teachers.

The selection of the items was based on the content objectives of the elementary science program. A determined effort was made to have the test reflect the structure of the Alberta science program. There are two physical science conceptual schemes, three biological science conceptual schemes, and one earth science conceptual scheme. Interwoven with these six content strands are objectives related to the methods of science and attitudes toward and knowledge of science and scientists.

At grades nine and twelve the science study is based on data gathered by using a standarized measure of student achievement, the Sequential Tests of Educational Progress (STEP), Series II. This test was accompanied by a measure of student's knowledge of and opinions about scientists and science as an area of study, the Test of Understanding Science (TOUS), form Jw. The choice of STEP II, form 3A for grade nine was made on the basis of its match with the curriculum, which was enhanced with the addition of 15 supplementary items. The choice of the TOUS, form Jw was in terms of the limited availability of tests measuring student achievement in this curricular area.

At the grade twelve level the emphasis was to be a more general knowledge of science, both in terms of the content of the high school program and of the more general aims and purposes of science. The choice of STEP II, form 2A was based on its relative quality and its suitability for gathering the information needed to complete the study.

In the following table the student performance is displayed as a percentage of correct responses for grade level and general content area. The number of items used in each area is also shown. The reader is again cautioned against making unwarranted comparisons across grades, since the tests are similar but have not been equated.

Table 1-14

SUMMARY OF PERCENTAGE OF CORRECT RESPONSES OF STUDENTS ON SCIENCE TESTS

	Grade 3 %	Grade 6 %	Grade 9 %	Grade 12 %
Physical Science	55.2 (21)*	59.6 (38)	62.6 (17)	55.5 (32)
Biological and Life Science	63.7 (26)	61.8 (45)	69.3 (12)	72.7 (16)
Earth Science	77.2 (6)	70.8 (11)	56.7 (17)	47.9 (3)
Methods of Science	57.4 (41)	64.9 (24)	62.4 (19)	64.8 (36)
Average Student Achievement in Content Areas	60.3	62.5	62.2	62.2
Science as a Human Endeavour	60.6 (14)	71.1 (14)	49.2 (45)	57.3 (45)
Interest in Science	66.1 (6)	61.2 (12)	---	---

* Number of test item is shown in parentheses.

1.8.1 Grade Three

Physical science, one of the more difficult conceptual areas, had the lowest student performance. Items dealing with concepts such as electricity and molecules and some basic energy conservation ideas were poorly answered.

Grade three students found the earth science items generally easier than those in other item groupings or clusters. The earth science concepts form a very minor portion of the curriculum but do capitalize on the interest that has been generated about stars and the solar system in general.

One of the major emphases of science programs in the past few years has been the area of the methods of science investigation. The relatively weak performance of third-grade students does not reflect this emphasis.

One surprise was the relative strength shown by grade three students in the area of science as a general field of human endeavour. These items were designed to check the students' perceptions of science and scientists, and the students tended to be reasonably knowledgeable about the field.

There are no differences in student achievement when sorted according to school size or location.

On items that attempted to measure student attitudes towards science, an average of 66.1 percent of the students responded favorably to questions related to their interest in science generally and some specific science topics in particular. The highest interest was indicated in responses to items asking about activities related to the natural environment.

1.8.2 Grade Six

Student performance on the physical science content items was not very high despite the increased emphasis on physical science in the curriculum at this level. There is an increased performance level relative to the life science items. In other words, the students found the life science items easier than the physical science ones, but the difference was not as great as with the grade three students.

The earth science area again proved to be an area of strength, even though it is not a major topic in the curriculum.

Stronger student performance on science methods items may reflect the increased importance that teachers and curricular supervisors have placed on this curricular area. Student perceptions of science and scientists appear to be relatively strong as well.

There are no significant differences in student achievement when the data are sorted according to school size or location.

In contrast to the grade three results, the average student performance was highest on those items related to the conservation of energy. The lowest student performance was on items in cluster five related to the change in environment and the response of organisms to change.

The grade six students showed a marked preference for science as compared with other areas of study (interest in science).

1.8.3 Grade Nine

The highest student performances in grade nine were on items dealing with life science, followed by physical science and earth science. The earth science performance was particularly weak and should be the subject of further investigation.

The area of science methods shows as an area of some strength. This may be the result of the increased emphasis on laboratory programs in junior high science.

The general area of science as a human activity, including such topics as aims of science and the skills and aptitudes of scientists, had a low response rate.

The boys responding to the test out-performed the girls significantly: $p < 0.05$ on the STEP II, which purports to measure knowledge and understanding of the fundamental concepts and processes of science. On the other hand, the girls out-performed the boys on the TOUS, form Jw, which is reported to measure knowledge and understanding about the scientific enterprise, about scientists and the tactics and methods of science.

The size and geographic location of the school did not prove to be a significant factor in differentiating student performance.

Of the four clusters of items into which the STEP II items were classified, the life science items as a group had the highest student achievement level and the earth science items had the lowest level. The grade nine course content items were handled second best, with scientific process items dropping to the bottom.

The TOUS items had lower average student performance levels. Since these items called for an opinion from the student and are in areas not often addressed directly by teachers, it might be expected that students would tend to have a greater diversity of responses.

On examination of the TOUS scores, all the differences are statistically significant. That is, the grade twelve students all scored better than those in grade nine, and at each grade level girls did better than boys.

The grade nine average student performance on the STEP II, form 3A was significantly superior ($p < 0.05$) to the 1970 U.S. grade nine student performance.

1.8.4 Grade Twelve

Biological science was the highest area of student performance. The lowest student performances were on earth science items.

The item cluster dealing with science methods showed a relatively strong performance, reflecting perhaps the emphasis that high school science courses place on science laboratory activities. Student performance on general knowledge items about the area of science as a field of human endeavour was relatively weak.

Grade twelve students reporting credit or enrolment in two 30-level sciences out-performed students in any other enrolment pattern including those taking three 30-level courses. Students reporting no science beyond the 20-level had a significantly lower average achievement level than any other group. It should be noted that the spread of scores was much greater for the group taking three sciences.

There were no significant differences ($p < 0.05$) when the student achievement scores were sorted according to school size.

The general level of achievement on the item clusters was highest on the biology items followed most closely by the item cluster dealing with the methods and processes of science.

The Alberta grade twelve average student achievement on the STEP II, form 2A was significantly better ($p < 0.05$) than the 1970 U.S. counterpart. That is, the general science knowledge of grade twelve students as measured on the Sequential Test of Educational Progress, Series II, form 2A was higher than for a similar group of grade twelve students.

1.8.5 Item Thought Level

The above-mentioned tests measure knowledge and understanding of the fundamental concepts and processes of science, comprehension and application of this knowledge, and mastery of science skills. Many of the items emphasize the use rather than the possession of knowledge.

The items can be categorized into three major areas: *Knowledge*, which measures the ability to recall ideas, materials or phenomena; *Comprehension*, which measures the ability to translate ideas or material from one method of expression to another and interpret material or extrapolate from it; and *Application*, which measures the ability to use learned information in answering an unfamiliar question or solving a new problem.

Table 1-15

AVERAGE STUDENT PERFORMANCE
ON CATEGORIES OF ITEMS (%)

Grade Level	Thought-Level Of Items		
	Knowledge	Comprehension	Application
3	68.3 (16)	57.8 (24)	67.0 (15)
6	67.4 (38)	57.7 (32)	59.3 (26)
9	64.7 (13)	69.6 (18)	58.6 (25)
12	61.5 (15)	63.1 (15)	58.6 (20)

Items dealing with the content areas of the school science program were categorized according to the thought level needed by the student in order to respond. Table 1-15 shows the average student performance for the various item categories. Relative to the knowledge base of the students, the average performance of elementary students on the comprehension items was lower than that of secondary students. Also relative to the knowledge level at each grade except grade three, the application scores are substantially lower. Grade three students answered application items almost as well as they answered knowledge items.

1.9 Conclusions

These data show the present level of student achievement in Alberta science programs on a number of dimensions. Whether this level is good, bad or average is very difficult to judge. The descriptive information is provided by this study to provide a basis for comparison, base-line data points for some future assessment. The value of this study will be determined at that future date.

The lower performance of the grade three students on the physical science items was thought to be acceptable because it

has been recognized that both teachers and students have experienced some difficulty with this area of the curriculum. The strong performances in both life science and earth-space science were judged to be very satisfying. It could be hypothesized that these student performances were influenced by the relative emphasis on the space theme by children's television programs and the efforts of such organizations as National Geographic in producing specials on the life science theme.

The only area in elementary science thought to be somewhat unsatisfactory was the student performance on items related to science methods. This poorer performance level is probably due to a different view of the role of student activity in science. Otherwise, student performance was generally satisfactory in the content dimensions of the program. However, the secondary student performance on the items related to the earth-space sciences was judged to be somewhat inadequate. One possible reason for this relatively poor student performance could be the level of abstraction of the concepts.

There is some concern about the adequacy of student performance on items related to the methods of scientists. Student performance levels at both grades nine and twelve were above the test average, so these were viewed as satisfactory. The program objectives related to this area are often neither accepted nor understood as being important. Another area of weak student performance was in understanding the nature of science, as tested by the Test of Understanding Science. This area is difficult to teach and is often viewed as being peripheral to the main intent of science programs.

In general terms, the student performances in grades three and six were satisfactory. Also in general terms, the student performances at grades nine and twelve were satisfactory with the exception of the junior high earth science area.

A further measure of adequacy was available at the secondary level because of the use of a standardized test with U.S. norms. The average student performance in Alberta at grades nine and twelve was above that of the 1970 U.S. norms. Since 1970 there has been a documented decline in standardized test scores across the U.S., so this difference suggests a much greater difference in performance could be shown if more recent norms were available. Using this as a standard, student performance in the field of secondary science in Alberta is satisfactory.

The investigation therefore concluded that, on balance, student achievement is satisfactory. A number of weak areas have been identified, and there are a number of areas in which student performance is quite strong.

The areas of some weakness include physical science in grades three and six, earth science in grade nine, and the grades nine the twelve responses to items asking about the purposes and aims of science as a field of human activity.

Areas in which there is some degree of satisfaction are those dealing with the methods and strategies of science and the life sciences.

1.10 Recommendation

The investigator recommends:

1.10.1 *that assessment of the science program in Alberta be a continuing process using the appropriate data from the present study as a base-line against which future achievement levels can be evaluated.*

The main value of undertaking an assessment of student achievement at this time lies in the use of the data as a point of comparison for future assessments. The present assessment efforts should represent a beginning point in the on-going evaluation of the education program in Alberta schools.

1.10.2 *that computerized item banking be developed and maintained.*

It is recognized that the instrumentation developed and/or purchased for this assessment was not wholly satisfactory and that further work on many of the items is needed to make them fully acceptable measures of achievement. The establishment of a computerized item bank would facilitate this process by simplifying access to the items and the accumulation of data about the items.

When an acceptable bank of items categorized according to program objectives is available, tests may easily be constructed to measure student achievement along a number of program dimensions. This data can then be accumulated to compile annual or biennial reports. By making the bank available to schools and teachers, the quality of testing in the province can be improved. The teacher gains access to a bank of proven items, and the banked items can be improved and extended by using input from teachers to develop and improve them.

Information about student performance on specific curricular objectives could be made available to curriculum developers on short notice. There could be the capability to respond very quickly to requests for information from those with a legitimate need.

1.10.3 *that a computerized bank of items used for the grades three and six testing be extended and improved.*

The items used at the elementary level were proven items from other assessment programs which measured achievement of a few selected objectives. Coverage of the program objectives should be broadened by the inclusion of more items. In general, the items were technically adequate but revision of some items could make them more applicable to the Alberta program.

1.10.4 *that computerized bank of items be developed for grades nine and twelve.*

To expedite the administration of the testing program, a commercially developed test was used to collect the base-line data. This provided some valuable insights into levels of student achievment in Alberta. But as with any instrument developed to sample the broad domain of program objectives, there are a few problems with making inferences about the quality of student achievement in Alberta. If an item bank were developed for Alberta, a better match between the test and the curriculum could be achieved.

A start on a bank of Alberta-valid items could be made by including items from former grade nine departmentals, other assessment programs, and locally developed tests.

The grade twelve item bank should focus less on the content dimension and more on the practical level of scientific and technical knowledge and applications that should be expected of all graduates of the Alberta education system.

1.10.5 *that the scope of the items in an item bank should continue to include the full spectrum of program objectives, including those in the affective domain.*

The exclusion of any particular program objectives may well imply that a lesser importance is placed on such objectives. It would be unfortunate if the weighting of curricular objectives came as a result of an unplanned and unconscious distribution of items. Student attitude towards science is another objective that should not be neglected.

1.10.6 *that standardized tests of educational progress should continue to be administered in order to make comparisons beyond provincial boundaries.*

It is too easy to become parochial and narrowly focussed in our view of program. To counter this tendency, a regular sampling of student achievement should be instituted as part of the continuing process of program evaluation.

1.10.7 *that the provincial testing program be conducted at a time of year which is most convenient.*

Separation of the provincial assessment program from school-based student evaluation procedures would serve to lessen the impact of over-testing in the year-end period. In situations where it is reasonable to return test results to the teacher, testing would take place at a time when the teacher can use such results to modify instruction.

1.10.8 *that an item validation procedure be instituted to capitalize on the lessons learned from the Results Interpretation Panels.*

The use of a broadly based reaction panel for the purpose of validating the content of the test items for use in either a bank or a provincially administered test is recommended. Further involvement of a community-based group of individuals would serve to widen the scope of provincial evaluations beyond the rather narrow view held by some curriculum specialists. If the aim of schooling is to produce an "educated" student, in the broader sense of community expectation, then representatives of that community should be involved in a reaction role at the initial steps of the evaluation and not just at the final stage.

1.10.9 *that a longitudinal study of a particular group of students (cohort) along a few broad curricular goals such as the scientific process be undertaken to investigate the grade placement of the specific objectives and their match with student competencies.*

Much of present science program has evolved as a result of experiences elsewhere and the availability of certain publishers' materials. It would be both logical and beneficial to gather information about the present nature of our science program before we expend resources on developing program materials or strategies to change the science program. The information gleaned from such a study would be of great benefit to the revision committees.

1.10.10 *that a further analyses of the data collected in this assessment be undertaken by qualified researchers either within Alberta Education or in the educational community.*

The anonymity of students and schools in the study would be fully protected, but there are many cross-correlations and factor analyses which were not made and which could in fact lead to an improved assessment.

1.10.11 *that a study of the effect of variables such as size of school, presence (or absence) of laboratories, and amount of time spent on science be commissioned by Alberta Education.*

The present study was limited in its scope by the very nature of the questions to be answered. However, this kind of process-product information is of value in examining some of the reasons for differences in student performance.

1.10.12 *that the testing procedures using matrix-sampling be continued in future province-wide assessments of student performance.*

The matrix-sampling procedure proved to be a very efficient, economical way of gathering valid data from across the province on a broad spectrum of objectives. One has to recognize that although a provincial testing program has little attraction for either teachers or students, the cooperation of these two groups is vital to its success. Therefore, limiting student time and minimizing the effort required of the teacher pays off in a greater degree of cooperation. From a provincial perspective, matrix sampling was facilitative. However, from a local perspective (if feedback is to be provided locally about individual pupils) it might not be applied. The purpose for the testing will dictate the appropriate sampling procedures.

PART II

RESULTS AND FINDINGS: A DETAILED EXAMINATION

1. INTRODUCTION

1.1 The Results of the Study

This section of the report presents the base-line data from the first administration of the MACOSA science tests with items keyed to the Alberta program of studies. The reader is cautioned that the item data provides only a first trial at assessing the level of student achievement. It is hazardous to make firm judgements about the quality of student achievement only on the basis of this data.

1.2 The Sample

The students responding to the science tests were enrolled in a 10 percent sample of Alberta schools in which grades three, six, nine and twelve were taught in the 1977-1978 school year.

In grades three and six the sampling was stratified by the location of the school and by the number of students enrolled in the target grade. The grades nine and twelve samples were stratified only by the number of students enrolled in the target grade.

Tables 1-1 and 1-2 show the number of schools in each cell of the sampling scheme. The number of schools in some categories was very small.

Table 1-1

ELEMENTARY SCHOOLS SELECTED BY SIZE AND LOCATION

	Enrolment	Large Urban	Small Urban	Rural	Total
Grade 3	<20	7	1	29	37
	21-30	9	3	11	23
	31-60	15	2	9	26
	>60	8	1	6	15
		39	7	55	101
Grade 6	<20	6	1	26	33
	21-30	7	2	12	21
	31-60	17	2	9	28
	>60	10	2	2	14
		40	7	49	96

Table 1-2

SECONDARY SCHOOLS SELECTED BY SIZE

	Enrolment	Schools		Enrolment	Schools
Grade 9	<20	11	Grade 12	<30	10
	21-30	3		31-60	5
	31-60	11		61-100	4
	61-100	5		>100	5
	101-150	5			24
	>150	5			
		40			

The students tested in these schools represented between 5 percent and 10 percent of the total population. There were 3073 grade three students, 2935 grade six students, 2426 grade nine students and 2125 grade twelve students. These numbers represent about 8 percent of the total number of students enrolled in these grades.

The sampling unit for this assessment is the school, and the reader is cautioned to interpret the resulting differences with this in mind. The decision to sample by schools was made to expedite the administration of the tests to identifiable groups of students.

1.3 The Tests

The primary level test was composed of items from a variety of sources; the largest single source was the National Assessment of Educational Progress. The NAEP-released items were an invaluable source of proven items. The curriculum outline as it existed in 1977-1978 for grades one, two and three was examined and an attempt was made to insure that each of the objective areas was measured by at least five items.

This resulted in 114 items on the primary test. The decision was made to incorporate a modified item-examinee sampling technique which would enable the tester to reduce the number of items any one student would write to less than 30. So that the results could be generalized to some degree, schools were randomly assigned to administer a random grouping of items. Because of this, it is not valid to make any comparison between schools or between individual students who did not write the same cluster of items. The test statistics reported are the pooled results for the sample as a whole; that is, for all 3073 students in the sample of 101 schools.

The items selected for administration were administered to a pilot sample in March 1978 and amended on the basis of the pilot results. The piloting of items provided a number of insights that led to procedural modifications as well. The grade three test was read to the students to reduce the effect of the reading level of the items, and in a number of cases the tests were distributed by express and courier services to improve the delivery system. A similar distribution procedure was followed with the elementary level test. There were 146 items selected for inclusion in the test given to grade six. Again a matrix sampling technique of items and examinees was used to reduce the number of items that a student would have to answer. The test statistics are the pooled results for the sample as a whole; that is, for all 2935 students in the sample of 96 schools. Forty-seven items are common to both the primary and elementary tests. This provides a small measure of comparison between the two tests.

A survey of the commercial tests available for junior high school science and the shortage of time for developing an item bank influenced the steering committee to decide that a combination of the Sequential Test of Education Progress, Series II, form 3A; the Test of Understanding Science; and a supplementary 15 items would serve to provide a general overview of student achievement on the seven objective clusters identified from the Alberta Program of Studies for junior high schools. This resulted in a test of 110 items which was judged to be manageable by the entire sample of junior high school students. The test statistics reported are for the whole group of 2426 students in 40 schools.

A similar decision was made for the senior high school test. The steering committee recommended that a survey of the general science competence of grade twelve students be made with a combination of the Sequential Test of Educational Progress, Series II, form 2A; the Test of Understanding Science, form Jw, and 12 supplementary items. This resulted in a test of 132 items which was judged to be manageable by the entire sample of 2125 students in 24 schools. The test statistics are reported for the entire sample and on the seven objective clusters.

1.4 The Statistics

An attempt has been made to report only those statistics which are meaningful and easily understood.

The student performance on each item is reported as the percentage of students answering the item correctly or, in some cases, the percentage of students responding to each foil or distracter. This "p-value" is also termed the "item difficulty" by some authors. The term p-value seems to better express student performance on test items for the purposes of this study.

The average student performance on various groupings of items or on entire tests is the mean of the "p-values." On an item cluster this mean is an index of student performance on the items grouped in the cluster.

The standard deviation is an indication of the spread of the scores, or the amount of the deviation of individual scores from the average or mean. It is a way of describing the degree of variation of scores for each group.

In comparing the means of various groups of students and the results on various clusters of items a simple test using students' "t" was used. For the calculations of the index and the distribution tables, the reader is referred to any elementary statistics text. Suffice it to say that any differences in means are statistically significantly different when they are sufficiently greater than the standard differences. That is, the probability is low (less than 0.05, 0.01, etc.) that the differences are due to chance factors.

Whenever tests of significance were made, two-tailed tests were used. This conservative approach to significance was used because of the dearth of experimental evidence that would enable one to predict differences in a given direction.

2. PRIMARY SCHOOL TEST RESULTS

2.1 Test Characteristics

The total of usable responses (N) from grade three students was 3073 in 101 schools.

The test consisted of 114 items divided randomly into four sub-tests. The test was administered by the classroom teacher reading the items to all grade three students registered in the selected schools and in attendence on May 17, 1978. The total test had a mean score of 57.5 percent with a standard deviation of 10 percent.

Table 2-1

AVERAGE GRADE 3 STUDENT RESPONSE TO THE TEST

$N = 3073$

$\overline{X} = 59.9\%$

2.2 Common Items

There were 47 items in common between the grade six test and the grade three test. Twenty of the items were targeted at grade three or lower, eleven were aimed at both levels, and fifteen were targeted at grade six. The grade six students, as one would expect, did significantly better ($p < 0.01$) on these items than did the grade three students.

Table 2-2

AVERAGE STUDENT PERFORMANCE ON THE COMMON ITEMS

	Grade 3	Grade 6
N	= 3073	2935
$\overline{X}$	= 57.4%	69.4%

2.3 Size and Location of Schools as Test Variables

The data provided in Table 2-3 shows that student achievement was relatively uniform across the province and among schools of varying class size and in different-sized systems across the province. None of the differences are significant for the size of the sample ($p > 0.01$)

Table 2-3

AVERAGE STUDENT PERFORMANCE ON THE
PRIMARY SCIENCE TEST

Size of School

# Of Students In Grade 3	≤20	21-30	31-60	>60	Total
N	37	23	26	15	101
$\overline{X}$ (%)	61.9	61.4	58.2	60.6	59.9
SD (%)	12.9	18.1	13.9	11.9	0.0

Location of School

	Urban Centre	Small Urban Centre	Rural	Total
N	39	7	55	101
$\overline{X}$ (%)	58.9	59.9	60.9	59.9
SD (%)	13.3	11.8	12.9	10.0

2.4 Grade Three Objective-Based Results

The criteria used for sorting the grade three item data were the program objectives taken from the Alberta Program of Studies for Elementary Schools--1975, p. 40-44. Each item on the test was matched with a group of related program objectives. The ten item clusters that resulted were:

1. Knowledge and understanding of the conceptual scheme A: When energy changes from one form to another, the total amount of energy remains unchanged.

2. Knowledge and understanding of the conceptual scheme B: When matter changes from one form to another, the total amount of matter remains unchanged.

3. Knowledge and understanding of the conceptual scheme C: Living things are interdependent with one another and with their environment.

4. Knowledge and understanding of the conceptual scheme D: A living thing is the product of its heredity and environment.

5. Knowledge and understanding of the conceptual scheme E: Living things are in constant change.

6. Knowledge and understanding of the conceptual scheme F: The universe and its components bodies are constantly changing.

7. Knowledge and understanding about scientists as people and as researchers.

8. Knowledge and understanding about the methods of science.

9. Knowledge and understanding about the scientific enterprise.

10. Interest in and value placed on the various fields of science.

For each cluster some sample items have been identified, the average student response to the items in the cluster has been reported, and the average student response for the item cluster (with the standard deviation of the item responses) have been calculated.

Within each cluster the items have been classified according to the portion of the objective cluster they appear to measure.

Table 2-4

GRADE 3 TEST RESULTS ON TEN ITEM CLUSTERS

Item Cluster	Average	# Of Items	Ranking
1. Energy	53.2	11	9
2. Matter	57.3	10	8
3. Interdependence	63.6	9	5
4. Heredity & Environment	66.0	12	3
5. Living Change	58.2	5	6
6. Earth-Space Science	77.2	6	1
7. Scientists	65.5	9	4
8. Scientific Methods	57.4	41	7
9. Scientific Enterprise	51.7	5	10
10. Interest In Science	66.1	6	2
		114	

The highest average student performances were on items in the sixth cluster. These items measured the earth-space science content of the program, a cluster which tends to be very popular with both teachers and students. The lowest average performances were in areas that are difficult and have available very little material suitable for use with young children. The results from Item Cluster #10, although included with the analysis, reflect the proportion of positive opinions held toward science; hence, the data are conceptually different from the data of the other clusters.

In the following sections, 2.4.1 to 2.4.10, item data from each cluster are presented independently.

2.4.1 Item Cluster #1

Knowledge and understanding of the conceptual scheme A: When energy changes from one form to another, the total amount of energy remains unchanged.

GRADE ONE:

1. Energy must be used to set an object in motion or to alter its motion.
2. Energy is used to do work.
3. Work is force acting through a distance.
4. Force is used to counteract force.

GRADE TWO:

1. The sun is our prime source of energy.
2. Chemical energy can be changed to light energy and heat energy.
3. Energy can be transferred from one place to another.
4. Energy can be transferred through the molecules of solids, liquids and gases.
5. Sound is the transfer of energy through the molecules of solids, liquids or gases.
6. Sounds vary in pitch; they may be high or low.
7. Sound travels through solids, liquids, or gases.
8. Sound results from the vibrations of molecules in solids, liquids or gases.
9. Sound waves travel through molecules of solids, liquids or gases.
10. Sound is a transfer of energy in a wave pattern through molecules of solids, liquids and gases.
11. Light is a form of energy.
12. Matter on the sun is converted to energy, including light energy.
13. Light is a form of energy transferred as a wave.
14. Sight is a physiological response to the stimulus of light energy.

GRADE THREE:

1. The sun is the earth's chief source of energy.
2. Energy can change from one form to another.

Table 2-5

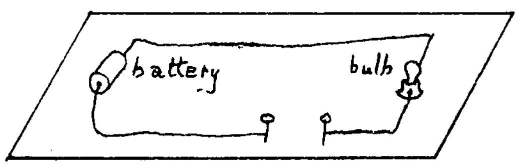
AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #1

Content Area of Items	# Of Items	Student Average (%)
Light as a form of energy	2	68.9
Sound as a form of energy	1	74.1
Electricity as a form of energy	5	51.6
Heat as a result of energy	2	40.3
Chemical change as a result of energy	1	34.3
	11	

Item cluster response - 4.8/11
Average student response to all items - 53.2%
Item cluster standard deviation - 27.4%

Sample Items - #219, 226

219. A boy was building a project in which a battery was to light a light bulb. But he did not have enough wire to finish his project and make the bulb light up.



What could he use to complete the project and make the bulb light up?

a. a wooden pencil
b. a piece of glass
c. a piece of plastic
*d. a paper clip

226. Two students made a string telephone from one house to the other. The sound of one person talking travels

a. through the air.
b. along the ground.
c. through the window.
*d. along the string.

The average student performance was 4.8 out of the 11 items, or 53.2 percent. As a group of items, the three dealing with electric circuits were handled most poorly. This is not altogether unexpected, since the concepts involved tend to be very abstract. Also, although they are included in the present program, they tend to be very difficult to teach. The items that students found easiest dealt with concepts with which students had had some out-of-school experiences such as rainbows, temperature, etc. In terms of the overall test, student performance ranked this cluster ninth out of ten.

2.4.2 Item Cluster #2

Knowledge and understanding of the conceptual scheme B: When matter changes from one form to another, the total amount of matter remains unchanged.

GRADE ONE:

1. Matter exists in various forms and states --solids, liquids and gases.
2. Heat may cause a change in the state of matter.
3. Evaporation and condensation are changes in the state of matter.

GRADE TWO:

1. A molecule is the smallest part of a substance which retains the chemical properties of that substance.
2. Heat energy causes water to expand.
3. Heat energy causes air to expand.
4. Heat energy causes matter to expand.

GRADE THREE: Matter consists of atoms and molecules.

Table 2-6

AVERAGE GRADE 3 STUDENT PERFORMANCE ON ITEM CLUSTER #2

Content Area Of Items	# Of Items	Student Average (%)
Nature of matter	2	53.9
Chemical change	2	69.3
Physical change	4	47.7
Conservation of mass	2	67.9
	10	

Item cluster response - 5.7/10
Average student response to all items - 57.3%
Item cluster standard deviation - 17.2%

Sample Items - #17, 50, 222

17. Iron is most likely to rust when it is

*a. damp.
b. dry.
c. painted.
d. covered with grease.

50. Which of the following will speed up the burning of a campfire?

*a. Blow on the fire.
b. Cover the fire with sand.
c. Sprinkle dirt on the fire.
d. Sprinkle water on the fire.

222. A jar is filled with ice. Then the lid is screwed on tight. After a few minutes the outside of the jar is wet. There are drops of water on the outside of the jar. This is evidence that the

a. lid is not on tight enough; the jar leaks.
b. jar was not dried last time it was washed.
c. jar is getting hot; it is sweating.
*d. moisture in the air is forming (condensing) on the jar.

The average student performance was 5.7 out of 10 items or 57.3 percent mastery of this objective cluster. In terms of the overall test, student performance ranked this cluster eighth out of ten. It should be noted that the content area of Clusters #1 and #2 is physical science and that these concepts tend to be more abstract and less observable than those in the life science and earth science areas.

2.4.3 Item Cluster #3

Knowledge and understanding of the conceptual scheme C: Living things are interdependent with one another and with their environment.

GRADE ONE:
1. There is an interchange of matter and energy between living things and their environment.
2. Organisms (living things) reproduce their own kind.
3. There is an interchange of matter and energy between living things and their environment. Adequate amounts of both are required for optimum growth.

GRADE TWO: Living things depend for their energy on a flow of materials from the environment.

GRADE THREE: There are characteristic environments, each with their characteristic life.

Table 2-7

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #3

Content Area Of Items	# Of Items	Student Average (%)
Characteristics of organisms	4	55.2
Characteristics of environments	2	71.0
Dependence of organisms on environment	3	69.9
	9	

Item cluster response - 5.7/9
Average student response to all items - 63.6%
Item cluster standard deviation - 22.4%

Sample Items - #13, 14

13. Which one of the following animals has scales over most of its body?

a. Bird
b. Frog
*c. Fish
d. Salamander

14. Which of the following animals usually hunts for food at night?

*a. Bat
b. Dog
c. Horse
d. Hummingbird

The average student performance was 5.7 out of 9 items for a 63.6 percent mastery of the item cluster.

In terms of the test as a whole, student performance ranked this cluster fifth out of ten.

2.4.4 Item Cluster #4

Knowledge and understanding of the conceptual scheme D: A living thing is the product of its heredity and environment.

GRADE ONE:
1. Organisms (living things) reproduce their own kind.
2. There is an interchange of matter and energy between living things and their environment. Adequate amounts of both are required for optimum growth. Size and structure are determined by heredity and environment.

GRADE TWO:
1. An organism is a product of its heredity.
2. The life and growth of a plant is affected by its environment.
3. An organism is a product of its heredity and environment.
4. There is an interchange of material and energy between organisms and their environment.

GRADE THREE:

Living things are related through possession of common structures.

Table 2-8

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #4

Content Area Of Items	# Of Items	Student Average (%)
Reproduction in organisms	4	59.9
Organic response to environment	4	71.3
Biological classification	4	66.8
	12	

Item cluster response - 7.9/12
Average student response to items - 66.0%
Item cluster standard deviation - 18.9%

Sample Items - #4, 19

4. Sometimes seeds stick to animals and are carried to new places where they will later grow. Which of these would most likely be spread this way?

a. Maple seed
b. Bean pod
c. Acorn
d. Cocklebur

19. Which of the following animals do NOT lay eggs?

a. Chickens
b. Whales
c. Frogs
d. Turtles

The average student performance was 7.9 out of 12 items, or 66 percent mastery of this item cluster.

In terms of the overall test, student performance ranked this cluster third out of ten. The concepts in this cluster (heredity and genetics) tend to be a little more difficult than those in Cluster #3.

2.4.5 Item Cluster #5

Knowledge and understanding of the conceptual scheme E: Living things are in constant change.

GRADE ONE: Animals of the past were different from the animals of the present.

GRADE TWO: Plants and animals have changed over the years.

GRADE THREE: Living things grow and develop in different environments.

Table 2-9

AVERAGE GRADE 3 STUDENT PERFORMANCE ON ITEM CLUSTER #5

Content Area Of Items	# Of Items	Student Average (%)
Fossils and environmental change	3	61.7
Growth patterns	2	53.0
	5	

Item cluster response - 2.9/5
Average student response to the items - 58.2%
Item cluster standard deviation - 17.1%

Sample Item - #23

23. The fossils of sea animals that lived long ago can be found in the

a. ocean water.
b. ice on a pond.
c. trunks of trees.
*d. rocks in the ground.

The average student performance was 2.9 out of 5 items, for a 58.2 percent mastery of the item cluster.

In terms of the test as a whole, student performance ranked this cluster as sixth out of ten.

2.4.6 Item Cluster #6

Knowledge and understanding of the conceptual scheme F: The universe, and its component bodies are constantly changing.

GRADE ONE: 1. The sun is the source of our light energy.

GRADE TWO:
1. Bodies in space are in constant motion.
2. The sun is the chief source of the earth's light.
3. Matter on the sun (and other stars) is converted.
4. Energy, including light energy.
5. The universe is constantly changing; its bodies are in constant motion.

GRADE THREE: There are seasonal and annual changes within the solar system.

Table 2-10

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #6

Content Area Of Items	# Of Items	Student Average (%)
Sun and stars	3	82.9
Earth	3	71.5
	6	

Item cluster response - 4.6/6
Average student response to the items - 77.2%
Item cluster standard deviation - 12.2%

Sample Item - #21

21. How many stars are there in the universe?

a. 12
b. 893
c. About a million
*d. More than have been counted

The average student performance was 4.6 out of 6 items for a 77.2 percent mastery of this item cluster. In terms of the

test as a whole, student performance ranked this cluster as first out of ten. This cluster measures concepts of earth and space science, an area which has an appeal for youngsters.

2.4.7 Item Cluster #7

Knowledge and understanding about scientists:

1. Generalizations about scientists as people
2. Abilities and skills needed by scientists

Table 2-11

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #7

Content Area Of Items	# Of Items	Student Average (%)
Skills and tasks	9	65.5

Item cluster response - 5.9/9
Average student response to the items - 65.5%
Item cluster standard deviation - 26.4%

Sample Item - #48

48. If you saw many scientists at work, a number of them would be

 a. selling soap.
 b. cooking meals.
 c. painting walls.
 *d. doing experiments.

The average student response to these items was 5.9 out of 9 items or 65.5 percent agreement with the preferred responses.

In terms of the total test, the student response pattern ranks this fourth of the ten clusters.

2.4.8 Item Cluster #8

Knowledge and understanding about the methods of science:

1. Generalities about the methods of science
2. Knowledge of the tactics and strategies of experimenting
3. Role of theories and models in science
4. Accumulation and verification of evidence
5. Relationship between science and technology

Table 2-12

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #8

Content Area Of Items	# Of Items	Student Average (%)
Observing	11	68.9
Classifying	7	60.0
Manipulation of data	7	60.8
Experimenting	1	33.6
Measuring	8	57.1
Inferring	7	37.0
	41	

Item cluster response - 23.5/41
Average student response to the items - 57.4%
Item cluster standard deviation - 21.8%

Sample Items - #8, 91, 98

8. You could measure the distance from your home to school in

a. dozens.
*b. kilometres.
c. kilograms.
d. litres.

91. If you want to find out how much a person grew in one year, which of the following <u>must</u> you know about the person?

a. His age
b. The type of food he eats
*c. His height at the start of the year
d. The height of his mother and father

98. John has a flat tire on his bicycle. He pumps the tire up with an air pump and begins to ride. In a few minutes the tire is flat again. To fix his tire, John must find

a. a better air pump.
b. whether the tire is made of rubber.
*c. where the air leaks out of the tire.
d. how many minutes it take the tire to go flat.

Many of the items that dealt with making observations and the identification of variables were handled very well. The items that dealt with the more complex skills of predicting, inferring and drawing conclusions were less well done. On this cluster, the average student response was 23.5 out of the 41 items for a 57.4 percent mastery, with a range from 68.9 percent in the observing items to 33.6 percent on the experimenting item. On the test as a whole, student performance ranked this cluster seventh out of ten.

2.4.9 Item Cluster #9

Knowledge and understanding about the scientific enterprise:

1. The human element in science
2. Communication among scientists
3. Role and limitations of instruments and procedures
4. Interaction of science and society
5. Purposes and aims of science
6. Unity and interdependence of science

Table 2-13

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #9

Content Area Of Items	# Of Items	Student Average (%)
Limitations of science	2	31.3
Purposes of science	2	66.1
Human element	1	63.6
	5	

Item cluster response - 2.6/5
Average student response to the items - 51.7%
Item cluster standard deviation - 19.5%

Sample Item - #104

104. In science one is LEAST likely to do which of the following things with an apple?

a. Find its mass.
b. Describe its color.
*c. Write a song about it.
d. Find out how many seeds it has.

The average student response to this item cluster was 2.6 out of 5 for a 51.7 percent agreement with the preferred responses. Knowledge of these concepts ranked tenth out of the ten clusters.

2.4.10 Item Cluster #10

Interest in and value placed on the various fields of science

Table 2-14

AVERAGE GRADE 3 STUDENT PERFORMANCE
ON ITEM CLUSTER #10

Content Area Of Items	# Of Items	Student Average (%)
Interest in science generally	2	71.4
Interest in specific areas	4	63.5
	6	

Item cluster response - 4.0/6
Average student response to the item - 66.1%
Item cluster standard deviation - 13.7%

Sample Item - #105

105. Do you find science topics interesting?

a. often
b. sometimes
c. never

Interest in science was measured in terms of items which asked students specifically which they preferred to study, read, watch or do--science activities or various arts, social studies or craft activities. On this basis, the average student response was 4.0 out of 6 items for a 66.1 percent agreement with the science-oriented response. Overall, student performance ranked this item cluster second of the ten.

3. UPPER ELEMENTARY SCHOOL TEST RESULTS

3.1 Test Characteristics

The total of usable student responses (N) from grade six students was 2935 in 96 schools.

The test consisted of 144 items randomly assigned to four sub-tests. The test was administered to all grade six students registered in the selected schools and in attendence on May 17, 1978. The total test had a mean of 63.6 percent.

Table 3-1

AVERAGE GRADE 6 STUDENT RESPONSE
TO THE TEST

N = 2935

$\overline{X}$ = 63.6%

3.2 Common Items

There were 47 items in common between the grade six test and the grade three test. Eighteen were targeted at grade four and higher, eleven at both levels, and eighteen at grade three and lower. The grade six students, as one would expect, did significantly better ($p < 0.01$) on these items than did the grade three students.

Table 3-2

AVERAGE STUDENT PERFORMANCE ON COMMON ITEMS

	Grade 3	Grade 6
N	3073	2935
$\overline{X}$	57.4%	69.4%

3.3 Size and Location of Schools as Test Variable

The data provided in Table 3-3 shows that student achievement was relatively uniform across the province and among schools of varying sizes of classes. None of these differences are significant for the size of the sample ($p > 0.01$).

Table 3-3

AVERAGE STUDENT PERFORMANCE ON THE ELEMENTARY SCIENCE TEST

Size of School
(# of students in Grade 6)

	≤20	21-30	31-60	>60	Total
N	33	21	28	14	96
$\overline{X}$	62.0%	65.2%	64.8%	62.1%	63.6%
SD	0.139	0.122	0.137	0.126	0.130

Differences are not significant ($p > 0.01$)

Location of School

	Large Urban	Small Urban	Rural	Total
N	40	7	49	96
$\overline{X}$	62.8%	66.5%	63.3%	63.6%
SD	0.128	0.123	0.134	0.130

Differences are not significant ($p > 0.01$)

3.4 Grade Six Objective-Based Results

The criteria used for sorting the grade six item data were the program objectives taken from the Alberta Program of Studies for Elementary Schools--1975, p. 45-47. Each item on the test was matched with a group of related program objectives. The ten item clusters that resulted were:

1. Knowledge and understanding of the conceptual scheme A: When energy changes from one form to another, the total amount of energy remains unchanged.

2. Knowledge and understanding of the conceptual scheme B: When matter changes from one form to another, the total amount of matter remains unchanged.

3. Knowledge and understanding of the conceptual scheme C: Living things are interdependent with one another and with their environment.

4. Knowledge and understanding of the conceptual scheme D: A living thing is the product of its heredity and environment.

5. Knowledge and understanding of the conceptual scheme E: Living things are in constant change.

6. Knowledge and understanding of the conceptual scheme F: The universe and its components bodies are constantly changing.

7. Knowledge and understanding about scientists as people and as researchers.

8. Knowledge and understanding about the methods of science.

9. Knowledge and understanding about the scientific enterprise.

10. Interest in and value placed on the various fields of science.

For each cluster some sample items have been identified, the average student response to the items in the cluster has been reported, and the average student response for the item cluster, with the standard deviation of the item responses, have been calculated.

Within each cluster the items have been classified according to the portion of the objective cluster they appear to measure.

Table 3-4

GRADE 6 TEST RESULTS ON
TEN ITEM CLUSTERS

Item Cluster		Average (%)	# Of Items	Ranking
1.	Energy	60.0	24	9
2.	Matter	58.8	14	8
3.	Interdependence	66.5	20	4
4.	Heredity & Environment	65.0	15	5
5.	Living Change	47.8	10	10
6.	Earth-Space Science	70.8	11	2
7.	Scientists	73.0	7	1
8.	Scientific Methods	65.8	24	5
9.	Scientific Enterprise	69.2	7	3
10.	Interest In Science	61.2	12	7

The highest average student performance was in Item Cluster #1, which represents a large portion of the grade six course, followed closely by the items dealing with scientists. The lowest performances were recorded in the clusters dealing with the conservation of matter and genetic changes, both of which are very difficult conceptually and in many schools more closely identified with the grade five program. The results from Cluster #10, although included for reference, reflect the proportion of positive attitudes toward science and science topics; hence the data are conceptually different from the data of other clusters.

In the following sections, 3.4.1 to 3.4.10, the data from each item cluster are presented independently.

3.4.1 Item Cluster #1

Knowledge and understanding of the conceptual scheme A: When energy changes from one form to another, the total amount of energy remains unchanged.

GRADE FOUR: A loss or gain of energy affects molecular motion.

GRADE FIVE: Energy must be applied to produce an unbalanced force resulting in motion or a change in motion.

GRADE SIX: The amount of energy obtained from a machine does not exceed the energy put into it.

Table 3-5

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #1

Content Area Of Items	# Of Items	Student Average (%)
Conservation of energy	4	66.6
Light as a form of energy	3	64.7
Heat and thermometry	5	47.9
Electricity and magnetism	4	62.0
Machines	8	61.6
	24	

Item cluster response - 14.4/24
Average student response to all items - 60.0%
Item cluster standard deviation - 19.6%

Sample Items - #68, 147, 150

68. It would take the most work to stop which of the following cars?

a. A light car going 30 km/h
b. A light car going 60 km/h
c. A heavy car going 30 km/h
*d. A heavy car going 60 km/h

147. Ball bearings are used in roller skates to

*a. reduce friction.
b. make the skates lighter.
c. reduce static electricity.
d. make them easier to oil.

150. Which of these can you not see through?

a. Air
b. Clear glass
c. Tinted glass
*d. Shiny white paper

The average student performance was 14.4 out of 24 items or 60 percent. The items with the poorest responses were those dealing with heat measurement. Removing these items from the test would have produced a student performance of 13.3 out of 20 items, or 66.3 percent mastery of the objective cluster.

In terms of the overall test, student performance ranked this cluster eighth out of ten.

3.4.2 Item Cluster #2

Knowledge and understanding of the conceptual scheme B: When matter changes from one form to another, the total amount of matter remains unchanged.

GRADE FOUR: In chemical change, atoms react to produce a change in the molecules.

GRADE FIVE: In chemical and physical change, the total amount of matter remains unchanged.

GRADE SIX: In nuclear reactions, a loss of matter is a gain in energy; and the sum of the matter and energy remains unchanged.

Table 3-6

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER # 2

Content Area Of Items	# Of Items	Student Average (%)
Nature of matter	5	55.0
Physical change	7	59.6
Chemical change	2	65.2
	14	

Item cluster response - 8.2/14
Average student response to all items - 58.8%
Item cluster standard deviation - 15.3%

Sample Items- #50, 51, 73

50. Which of the following will speed up the burning of a campfire?

*a. Blow on the fire.
b. Cover the fire with sand.
c. Sprinkle dirt on the fire.
d. Sprinkle water on the fire.

51. Which of the following dissolves LEAST in water?

*a. Glass
b. Salt
c. Soap
d. Sugar

73. To get salt out of salt-water, one could

a. cool the water.
b. add more water.
*c. boil away the water.
d. dissolve air in the water.

The average student performance was 8.2 out of 14 items or 58.8 percent mastery of this objective cluster. In terms of the overall test, student performance ranked this cluster ninth out of ten. The grade six students found Item Clusters #1 and 2, which deal with physical science, to be one of the most difficult areas of the curriculum.

3.4.3 Item Cluster #3

Knowledge and understanding of the conceptual scheme C: Living things are interdependent with one another and with their environment.

GRADE FOUR: Living things capture matter from the environment and return it to the environment.

GRADE FIVE: The capture of radiant energy by green plants is basic to the growth and maintenance of all living things.

GRADE SIX: Living things are adapted by structure and function to their environment.

Table 3-7

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #3

Content Area Of Items	# Of Items	Student Average (%)
Characteristics of organisms	6	68.0
Responses of organisms to their environment	9	75.9
Interdependence of organisms	5	47.8
	20	

Item cluster response - 13.3/20
Average student response to all items - 66.5%
Item cluster standard deviation - 19.9%

Sample Items - #13, 14

13. Which one of the following animals has scales over most of its body?

 a. Bird
 b. Frog
*c. Fish
 d. Salamander

14. Which of the following animals usually hunts for food at night?

*a. Bat
b. Dog
c. Horse
d. Hummingbird

The average student performance was 13.3 out of 20 items for a 66.5 percent mastery of the item cluster.

In terms of the test as a whole, student performance ranked this cluster fourth out of ten.

3.4.4 Item Cluster #4

Knowledge and understanding of the conceptual scheme D: A living thing is the product of its heredity and environment.

GRADE FOUR:
1. A living thing reproduces itself and develops in a given environment.
2. A living thing is the product of its heredity and environment.

GRADE FIVE: The cell is the unit of structure and function; a living thing develops from a single cell.

GRADE SIX: The characteristics of a living thing are laid down in a genetic code.

Table 3-8

AVERAGE GRADE 6 STUDENT PERFORMANCE ON ITEM CLUSTER #4

Content Area Of Items	# Of Items	Student Average (%)
Reproduction	4	73.1
Structure and function	9	62.7
Heredity and genetic code	2	59.5
	15	

Item cluster response - 9.8/15
Average student response to all items - 65.0%
Item cluster standard deviation - 13.8%

Sample Items - #19, 27

19. Which of the following animals do NOT lay eggs?

a. Chickens
*b. Whales
c. Frogs
d. Turtles

27. What is the most important thing that the lungs do?

a. Hold the chest out.
b. Protect against germs.
c. Pump the blood through the body.
*d. Provide a place for oxygen to enter the blood.

The average student performance was 9.8 out of 15 items, or 65 percent mastery of this item cluster.

In terms of the total test, student performance ranked this cluster sixth out of ten.

3.4.5 Item Cluster #5

Knowledge and understanding of the conceptual scheme E: Living things are in constant change.

GRADE FOUR: The environment is in constant change.

GRADE FIVE: Living things have changed over the ages.

GRADE SIX: Changes in the genetic code produce changes in living things.

Table 3-9

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #5

Content Area Of Items	# Of Items	Student Average (%)
Environment is in constant change	7	42.5
Living things have changed	3	60.0
	10	

Item cluster response - 4.8/10
Average student response to the items - 47.8%
Item cluster standard deviation - 18.6%

Sample Item - #23

23. The fossils of sea animals that lived long ago can be found in the

a. ocean water.
b. ice on a pond.
c. trunks of trees.
*d. rocks in the ground.

The average student performance was 4.8 out of 10 items for a 47.8 percent mastery of the item cluster. Seven of the ten items related to a weather map of Western Canada and the other three items related more specifically to changes in the environment over time. It may well be that the items have influenced student performance, which ranked this cluster as tenth out of ten.

3.4.6 Item Cluster #6

Knowledge and understanding of the conceptual scheme F: The universe, and its component bodies are constantly changing.

GRADE FOUR: The motion and path of celestial bodies are predictable.

GRADE FIVE: Bodies in space, as well as their matter and energy, are in constant change.

GRADE SIX: Nuclear reactions produce the radiant energy of stars, and variations in these reactions result in change.

Table 3-10

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #6

Content Area Of Items	# Of Items	Student Average (%)
Sun and stars	2	92.6
Solar system	9	66.0
	11	

Item cluster response - 7.8/11
Average student response to the item - 70.8%
Item cluster standard deviation - 15.5%

Sample Items - #21, 34

21. How many stars are there in the universe?

a. 12
b. 893
c. About a million
*d. More than have been counted

34. The only star you usually see in the daytime is

*a. the sun.
b. the moon.
c. the Pole Star.
d. Mars.

The average student performance was 7.8 out of 11 items for a 70.8 percent mastery of the item cluster. Overall, student performance ranked this cluster second out of ten.

3.4.7 Item Cluster #7

Knowledge and understanding about scientists

1. Generalizations about scientists as people
2. Abilities and skills needed by scientists

Table 3-11

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #7

Content Area Of Items	# Of Items	Student Average (%)
Skills and tasks	7	73.0

Item cluster response - 5.1/7
Average student response to the item - 73.0%
Item cluster standard deviation - 28.6%

Sample Item - #48

48. If you saw many scientists at work, a number of them would be

a. selling soap.
b. cooking meals.
c. painting walls.
*d. doing experiments.

The average student response was 5.1 of 7 items or 73 percent agreement with the preferred responses.

In terms of the total test, the student response pattern ranks this cluster as first of ten.

3.4.8 Item Cluster #8

Knowledge and understanding about the methods of science:

1. Generalities about the methods of science
2. Knowledge of the tactics and strategies of experimenting
3. Role of theories and models in science
4. Accumulation and verification of evidence
5. Relationship between science and technology

Table 3-12

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #8

Content Area Of Items	# Of Items	Student Average (%)
Identification of variable	1	95.8
Manipulation of data	5	75.5
Experimenting	3	71.7
Quantifying	9	78.0
Inferring	6	32.6

Item cluster response - 15.8/24
Average student response to the items - 65.8%
Item cluster standard deviation - 23.1%

Sample Items - #8, 91, 98

8. You could measure the distance from your home to school in

a. dozens.
*b. kilometres.
c. kilograms.
d. litres.

91. If you want to find out how much a person grew in one year, which of the following <u>must</u> you know about the person?

a. His age.
b. The type of food he eats.
*c. His height at the start of the year.
d. The height of his mother and father.

98. John has a flat tire on his bicycle. He pumps the tire up with an air pump and begins to ride. In a few minutes the tire is flat again. To fix his tire, John must find

a. a better air pump.
b. whether the tire is made of rubber.
*c. where the air leaks out of the tire.
d. how many minutes it take the tire to go flat.

Average student performance on the items in this cluster was much better than that of the primary students. As one would expect, the interpretation of graphs was substantially better, as were measuring and quantifying skills. The average student response was 15.8 out of 24 items or 65.8 percent mastery of this cluster. Overall, student performance on this item cluster ranked fifth out of ten.

3.4.9 Item Cluster #9

Knowledge and understanding about the scientific enterprise:

1. The human element in science
2. Communication among scientists
3. Role and limitations of instruments and procedures
4. Interaction of science and society
5. Purposes and aims of science
6. Unity and interdependence of science

Table 3-13

AVERAGE GRADE 6 STUDENT PERFORMANCE
ON ITEM CLUSTER #9

Content Area Of Items	# Of Items	Student Average (%)
Limitations of science	4	70.2
Purposes	4	68.0

Item cluster response - 4.8/7
Average student response to the items - 69.2%
Item cluster standard deviation - 8.7%

Sample Item - #103

103. When a scientist takes many measurements during an experiment and studies the results carefully what is he probably doing?

a. Changing a theory into a law.
b. Changing a law into a theory.
c. Trying to correct a mistake he has made in arithmetic.
*d. Seeking for relationships among the measurements he has obtained.

The average student response to this item cluster was 4.9 out of 7 items for a 69.7 percent agreement with the preferred responses. Knowledge about the scientific enterprise was third out of ten, second only to knowledge about scientists.

3.4.10 Item Cluster #10

Interest in and value placed on the various fields of science.

Table 3-14

AVERAGE GRADE 6 STUDENT PERFORMANCE ON ITEM CLUSTER #10

Content Area Of Items	# Of Items	Student Average (%)
Interest in science generally	6	73.8
Earth-space-physical biological	3	62.8
Value placed on science	3	34.4

Item cluster response - 7.3/12
Average student response to the item - 61.2%
Item cluster standard deviation - 21.4%

Sample Items - #105, 111

105. Do you find science topics interesting?

a. Often
b. Sometimes
c. Never

111. In the past month or so I have read a story or article about (choose the most interesting one)

a. space exploration or astronomy.
b. rocks and minerals or mining.
*c. animals or plant life.
d. scientists or medical doctors.
e. other non-science topics.

Interest in science was higher than the value placed on science. That is, this age group indicated a general interest in the area. Student response was 6.3 out of 9 items for a 70 percent agreement with the science-oriented response, as opposed a student response of 1.03 out of 3 items for a 34.4 percent agreement with the accepted response for the value placed on science. Overall, student performance ranked this cluster seventh out of ten.

4. JUNIOR HIGH SCHOOL SCIENCE TEST RESULTS

4.1 Test Characteristics

The total of usable responses (N) was 2426 in 40 schools.

The test consisted of three instruments (sub-tests): the 50-item Sequential Test of Educational Progress (STEP), a 15-item test focused on areas not covered by STEP, and the 45-item Test Of Understanding Science (TOUS), form Jw.

Table 4-1

GRADE 9 AVERAGE STUDENT RESPONSE

	Average Student Response (Raw Score)	SD
STEP II 3A	34.1	6.92
Supplement	6.3	2.42
TOUS	21.8	5.98
TOTAL	62.2	13.02

4.2 STEP II Results

On the STEP II test, the items with the lowest performance scores are #25, 35, 36, 37, 38, 40, 41 and 43, all of which are items calling for some application or synthesis of information.

Table 4-2 shows an item-by-item comparison of the 1978 proportion of students choosing the keyed response compared with the 1970 norms established by the test developers.

An examination of the item-by-item comparisons of the Alberta average student responses with the U.S. norming sample shows that on most items the Alberta average student performance was the same as or better than the performance of their American counterparts. On only seven items was the local performance lower than that of the U.S. norming sample.

4.3 TOUS Results

TOUS, form Jw, has 45 items that are grouped in three clusters, dealing with the scientific enterprise, scientists, and the methods and aims of science.

Table 4-2

STEP II Form 3A - Science
COMPARISON OF UNITED STATES AND ALBERTA GRADE NINE STUDENTS

		Percent Passing (Norms) Grade 9 -- % Right	
Item #	Right Answer	Alberta	U.S.A.
1	D	95	94
2	C	94	92
3	A	87	88
4	C	92	92
5	D	84	86
6	C	82	86
7	D	91	88
8	B	96	95
9	A	73	72
10	D	94	94
11	A	55	58
12	D	76	74
13	C	77	73
14	C	77	68
15	A	78	77
16	D	69	63
17	B	88	65
18	C	76	66
19	C	56	60
20	C	74	64
21	B	79	72
22	D	69	58
23	B	73	67
24	B	68	63
25	C	47	58
26	A	65	61
27	A	57	63
28	C	76	66
29	B	80	76
30	D	54	39
31	C	65	60
32	C	65	54
33	D	67	57
34	D	65	60
35	D	44	43
36	C	41	38
37	A	47	44
38	A	44	40
39	C	55	51
40	A	47	40
41	D	40	39
42	A	53	43
43	D	47	40
44	A	51	50
45	D	71	69
46	C	50	49
47	B	50	43
48	B	84	81
49	A	79	73
50	D	71	69

The instrument was developed by researchers at Harvard University Graduate School of Education to measure student learning in the so-called "intangible" areas of science. These intangibles include the nature of scientific inquiry, the role of scientists as inquirers, and their relationship to science as an institution in our technological society.

Table 4-3

GRADES 9 AND 12 STUDENT RESPONSES TO TOUS: RAW SCORE

	Grade 9			Grade 12		
	Boys	Girls	Total	Boys	Girls	Total
N	1234	1144	2426	1059	1034	2125
Average Raw Score	21.9	22.6	21.8	24.6	25.8	25.3
Standard Deviation	6.1	5.8	6.0	6.7	6.2	6.5

The data presented in Table 4-3 indicate that there is a growth of knowledge and understanding in this domain from grade nine to grade twelve. Whether the extent of the difference is enough cannot be answered within the framework of this study. However, the differences are significant ($p < 0.001$) given the number of students involved.

4.4 Sex Differences Among Grade Nine Students

There were 1234 males and 1144 females responding to the junior high tests, with 48 not recording the information.

Table 4-4

AVERAGE STUDENT RESPONSE TO
THE JUNIOR HIGH TESTS: RAW SCORE

N	Male 1234		Female 1144	
	$\overline{X}$	SD	$\overline{X}$	SD
STEP II	35.1	6.9	33.2	6.7
Supplement	6.7	2.5	5.9	2.3
TOUS	21.1	6.1	122.6	5.8
TOTAL	62.9	13.3	61.7	12.6

Table 4-5

TEST ON DIFFERENCES BETWEEN PAIRS OF MEANS
USING STUDENTS' t (GRADE 9)

			Male STEP	Male TOUS	Male TOTAL
			35.1	21.1	62.9
	STEP	33.2	***	-	-
Female	TOUS	22.6	-	***	-
	TOTAL	61.7	-	-	*

* t ($p \leq 0.05$) differences are "real" 95% of the time.
** t ($p \leq 0.01$) differences are "real" 99% of the time.
*** t ($p \leq 0.001$) differences are "real" 99.9% of the time.

Boys did significantly better than girls on the content questions while girls did significantly better than boys on the questions related to science and scientists.

4.5 Size of School as a Variable

The sample schools were classified by their reported grade nine enrolment. The average student performance on the TOUS and STEP tests is reported for each category of school. The total results follow the same pattern; the supplementary items did not influence the pattern.

Table 4-6

GRADE 9 STEP II MEAN RAW SCORES
BY SIZE OF SCHOOL

School Class Size	<20	21-30	31-60	61-100	101-150	>150	All
$\overline{X}$	34.9	35.3	34.1	35.1	34.4	32.1	34.1
SD	7.3	6.4	6.4	6.7	6.8	7.4	6.9
N	11	3	1	5	5	5	40

It is most noticeable that students in the largest junior high schools (more than 150 students) do less well than students in every other category of school. This difference is not significant in terms of the relatively small sample of schools in each category, but in practical terms, students in smaller schools (less than 150 students) appear to do better on the standardized tests of general science competency.

Table 4-7

GRADE 9 TOUS MEAN RAW SCORES
BY SIZE OF SCHOOL

School Class Size	<20	21-30	31-60	61-100	101-150	>150	All
$\overline{X}$	20.8	22.9	21.6	22.0	22.6	20.0	21.8
SD	6.8	6.	5.	5.5	5.9	6.0	6.0
N	11	3	11	5	5	5	40

Again, the students in the largest junior high schools have done only somewhat less well on this test than students from smaller schools. The differences are significant at the 0.2 level for these rather small samples of the schools in each category. This pattern may, however, indicate a focus for future studies.

4.6 Grade Nine Objective-Based Results

The referenced criteria were the program objectives taken from the Alberta Program of Studies for Junior High Schools--

1975, p. 25-29M. Each item on the three-part test has been matched with a cluster or group of related program objectives. The seven clusters are:

#1 Grade 7 course content - life science
#2 Grade 8 course content - earth science
#3 Grade 9 course content - physical science
#4 Science process skills (not graded)
#5 Understanding of the scientific enterprise
#6 Understanding of scientists
#7 Understanding about science and sciencing

For each cluster the number of items have been identified, the average student reponse to the item has been reported, and the average response for the cluster (with the standard deviation) has been calculated.

Within each of the content clusters the items have been classified according to the sub-objective to which they appear to relate. The classification of the items was based mainly on the publishers' manuals for the tests which were used to make up the junior high school test battery.

Table 4-8

GRADE 9 RESULTS BY ITEM CLUSTER

Item Cluster		Average (%)	Items	Rank
1.	Life Science	69.3	12	1
2.	Earth Science	56.7	17	5
3.	Physical Science	62.6	17	2
4.	Methods of Science	62.4	19	3
5.	Scientific Enterprise	51.6	8	6
6.	Scientists	58.7	11	4
7.	Sciencing	44.5	26	7

Item Cluster #1 represents a sample of the content from the grade seven science program. The concepts tested tend to be those that are reinforced by the individual student's daily or seasonal observations of the environment. It is to be expected, therefore, that student achievement is at a satisfactory level.

Item Cluster #2 represents a sample from the concepts taught in grade eight. It would seem that either these concepts are not of sufficient interest to be retained over the year or else they are not being reinforced through observation.

The items in Cluster #3 are from the content of the current year's course. The second place ranking indicates some degree of familiarity with the concepts.

The items in Cluster #4 test the procedures and methods learned in the laboratory and may be an indicator of the amount of experience students are getting in this area.

Clusters #5, 6 and 7 represent dimensions of the TOUS test and represent knowledge and understanding of the scientific enterprise as practised by members of the scientific community.

4.6.1 Item Cluster #1 - Grade 7 Course Content - Life Science

Table 4-9

AVERAGE GRADE 9 STUDENT PERFORMANCE
ON ITEM CLUSTER #1

Content Area Of Items	# Of Items	Student Average (%)
Classification, structure and function	3	78.5
Life processes	5	65.5
Interdependence of living things	4	67.0
	12	

Item cluster response - 8.3/12
Average student response - 69.3%
Item cluster standard deviation - 16.7%

Sample Items - #15, 18

15. Farmers keep hives of bees in orchards to supply honey and to

- • a. pollinate tree blossoms.
- b. provide food for the wild animals.
- c. maintain a balance of nectar.
- d. keep the soil fertile.

18. For quick energy before a race, it would be best for a runner to have

- a. a cup of coffee.
- b. a frankfurter.
- • c. a candy bar.
- d. hard-boiled eggs.

4.6.2 Item Cluster #2 - Grade 8 Course Content - Earth Science

Table 4-10

AVERAGE GRADE 9 STUDENT PERFORMANCE ON ITEM CLUSTER #2

Content Area Of Items	# Of Items	Student Average (%)
Space, solar system	3	64.0
Earth's weather and climate	6	41.2
Earth's surface, rocks and water	8	65.7
	17	

Item cluster response - 9.6/17
Average student response - 56.7%
Item cluster standard deviation - 24.5%

Sample Items - #13, 26, 51

13. The presence of sea fossils in the Rocky Mountains is explained by which of the following?

a. The mountains were formed by a huge volcano.
b. Sea animals migrated into the mountains.
• c. The region was once covered by an ocean.
d. The region was once crossed by glaciers.

26. Which of the following accounts for many spring floods in the United States and Canada?

• a. Snow may melt rapidy in the spring.
b. Cyclic upheavals of the molten core of the earth occur in the spring.
c. Water flows more rapidy in spring than in winter.
d. The moon is closer to the earth in spring.

51. Most of the energy to power the physical systems on our planet comes from

a. the moon.
b. inside the earth.
• c. the star that we call the sun.
d. outside the solar system.
e. the rotation of the earth.

4.6.3 Item Cluster #3 - Grade 9 Course Content - Physical Science

Table 4-11

AVERAGE GRADE 9 STUDENT PERFORMANCE ON ITEM CLUSTER #3

Content Area Of Items	# Of Items	Student Average (%)
Matter, space and mass	4	63.7
Kinetic Molecular theory	5	60.8
Conservation of matter and energy	5	52.9
Physical and chemical change	3	80.1
	17	

Item cluster response - 10.6/17
Average student response - 62.6%
Item cluster standard deviation - 15.8%

Sample Items - #7, 14, 17

7. A chemical change occurs when the atoms present in a material are rearranged or the molecules are changed enough to form something new. Which of the following is a chemical change?

a. Breaking glass
b. Cutting wood
c. Spilling milk
• d. Burning alcohol

14. Object X weighs twice as much as object Y when both objects are at the same place on earth. The mass of object X is

a. half that of object Y.
b. the same as that of object Y.
• c. twice that of object Y.
d. four times that of object Y.

17. The density of a sample is determined from its

a. size and color.
• b. mass and volume.
c. shape and volume.
d. size and shape.

4.6.4 Item Cluster #4 - Science Process Skills

Table 4-12

AVERAGE GRADE 9 STUDENT PERFORMANCE
ON ITEM CLUSTER #4

Content Area Of Items	# Of Items	Student Average (%)
Observing and classifying	1	68.8
Experimental procedure	2	65.0
Quantifying	1	36.8
Data manipulation	4	52.4
Data interpretation	7	77.0
Problem indentification	1	56.0
Hypothesizing, inferring and predicting	3	48.6
	19	

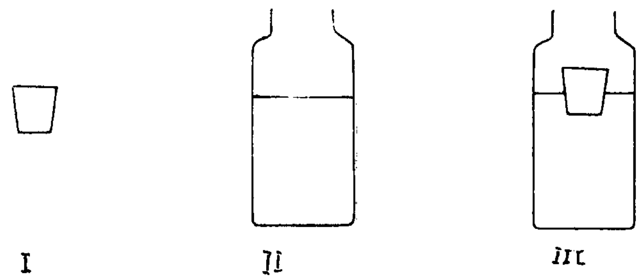
Item cluster response - 11.9/19
Average student response - 62.4%
Item cluster standard deviation - 19.2%

Sample Items - #16, 32, 40

16. Which of the following represents a scientific observation and not a superstition?

a. Bats commonly get tangled in hair.
b. Toads cause warts.
c. A black cat crossing one's path brings bad luck.
• d. If seeds are planted upside down, the stems grow upward.

32.



John weighed a cork (I). He then weighed a small jar of water (II). He put the cork in the jar of water and it floated. He weighed the jar of water with the cork floating in it (III). He found that III weighed the same as

a. I.
b. II.
• c. I + II.
d. II - I.

40. Ken claps his hands sharply to attract Annette's attention. He hears a distinct echo from the direction of the cliff. In order to hear the echo louder, he walks much closer to the cliff. He finds that he can no longer hear a distinct echo. Which of the following best explains this?

• a. The echo now reaches Ken so soon that he cannot tell the echo from the handclap.
b. The cliff is high and the sound is mostly reflected over Ken's head when he is closer to the cliff.
c. The sound of the echo becomes fainter as Ken gets closer to the cliff.
d. The closer Ken is to the cliff, the more the sound is absorbed by the cliff.

4.6.5 Item Cluster #5 - Understanding of the Scientific Enterprise

Table 4-13

AVERAGE GRADE 9 STUDENT PERFORMANCE
ON ITEM CLUSTER #5

Content Area Of Items	# Of Items	Student Average (%)
Communication among scientists	3	54.2
Instrumentation	1	47.9
Money and resources	1	47.7
International character	1	79.6
Interaction of science and society	2	37.5
	8	

Item cluster response - 4.1/8
Average student response - 51.6%
Item cluster standard deviation - 17.0%

Sample Items - #72, 101

72. The scientists of today can work on more complex problems than the scientists of the past mainly because they

a. work much harder than earlier scientists.
b. have more imagination than earlier scientists.
• c. build on the work of earlier scientists.
d. are more intelligent than earlier scientists.

101. A young person who wants to have a career in science would be able to find a job as a scientist

a. only in the United States or Russia.
b. only in countries in American and Europe.
• c. in most countries of the world.
d. only in countries with large industries.

4.6.6 Item Cluster #6 - Understanding of Scientists

Table 4-14

AVERAGE GRADE 9 STUDENT PERFORMANCE
ON ITEM CLUSTER #6

Content Area Of Items	# Of Items	Student Average (%)
Generalizations about scientists as people	5	64.4
Institutional pressure on scientists	1	46.8
Abilities needed by scientists	5	55.5
	11	

Item cluster response - 6.5/11
Average student response - 58.7%
Item cluster standard deviation - 12.6%

Sample Items - #82, 97

82. Mary writes down all the details of her science experiments. If Mary becomes a scientist, this training will help her to

a. be patient in doing her experiments.
• b. make better reports about her experiments.
c. think up theories from her experiments.
d. work out new experiments to perform.

97. In trying to decide whether or not to go to a new movie, a scientist would probably ask himself:

a. "Does the film use experimental techniques?"
b. "Is there cause and effect in the plot?"
• c. "Will I like this film?"
d. "Is the plot factual and accurate?"

4.6.7 Item Cluster #7 - Understandings About Science

Table 4-15

AVERAGE GRADE 9 STUDENT PERFORMANCE
ON ITEM CLUSTER #7

Content Area Of Items	# Of Items	Student Average (%)
Scientific methods: Tactics and strategies	9	36.8
Theories and models	7	52.6
Aims of science	6	48.2
Science and technology	2	32.9
Unity and interdependence of the sciences	2	51.0
	26	

Item cluster response - 11.6/26
Average student response - 44.5%
Item cluster standard deviation - 16.0%

Sample Items - #81, 98, 100

81. The chief purpose of the science of botany is to

a. find out what plants grow best in various kinds of soil.
• b. understand how plants live, grow, and reproduce.
c. develop new drugs and medicines from plants.
d. help farmers grow bigger plants and produce more food.

98. When we say that a scientist forms a hypothesis about an experiment, we mean that he

• a. makes a careful guess about what will happen.
b. gives directions for doing the experiment properly.
c. suggests how to make exact measurements.
d. describes how the experiment was carried out.

100. A scientific law describes

a. rules which scientists must obey.
• b. relationships between events in nature.
c. directions for doing good experiments
d. good guesses about how things happen.

5. SENIOR HIGH SCHOOL SCIENCE TEST RESULTS

5.1 Test Characteristics

The total number of usable responses (N) was 2125 in 24 schools, a 10 percent sample of senior high schools. The test consisted of a combination of three instruments: the 75-item Sequential Test of Educational Progress (STEP) Series II, form 2A for science, a 12-item supplement and the 45-item Test of Understanding Science (TOUS), form Jw.

Table 5-1

GRADE 12
AVERAGE STUDENT RESPONSE

	Raw Score $\overline{X}$	SD
STEP II, 2A	46.90	18.32
Supplement	6.74	2.03
TOUS	25.29	6.49
TOTAL	78.92	18.21

5.2 STEP Test Results

Items causing the greatest difficulty on this test were those linked specifically to knowledge taught in science courses (items #32 - optics, #40 - chemistry, #42 - rocks, #45 - space, #69 - electrostatics, #70 - graph, #71 - chemistry).

Similarly, the supplementary questions that gave the most difficulty to students were items #87, which dealt with the source of hydrocarbons, and #84, which dealt with a chemical equation.

The grade twelve students in Alberta outperformed their U.S. counterparts by 5 raw score points, a significant difference ($p < 0.05$). In addition, Table 5-2 shows that for 73 of the 75 items the Alberta students had the same or a higher proportion of passes.

5.3 TOUS Results

TOUS, form Jw, has 45 items that are grouped in three clusters, dealing with the scientific enterprise, scientists, and the methods and aims of science.

Table 5-2

STEP II Form 2A - Science
COMPARISON OF UNITED STATES AND ALBERTA GRADE TWELVE STUDENTS

		Percent Passing (Norms) Grade 12 -- % Right	
Item #	Right Answer	Alberta	U.S.A.
1	D	86.4	83
2	C	91.8	87
3	B	51.2	50
4	C	83.1	72
5	A	65.2	55
6	D	63.3	63
7	A	67.0	56
8	C	67.8	58
9	D	71.8	63
10	D	78.8	67
11	C	72.9	67
12	C	93.9	91
13	A	92.3	82
14	D	71.3	55
15	B	89.1	83
16	B	78.9	74
17	A	60.6	52
18	D	53.3	56
19	B	67.0	57
20	A	47.4	44
21	B	59.2	52
22	B	55.6	42
23	C	69.0	51
24	B	60.8	47
25	B	40.1	36
26	B	62.1	44
27	A	47.2	35
28	C	47.6	41
29	B	46.4	40
30	A	72.9	72
31	B	41.9	34
32	D	37.9	38
33	A	61.4	37
34	C	89.0	85
35	C	56.2	40
36	A	76.9	71
37	A	55.9	39
38	A	60.9	45
39	B	58.5	46
40	A	38.4	30
41	B	43.9	41
42	B	32.8	35
43	D	47.8	32
44	D	46.4	42
45	D	34.4	31
46	B	92.5	86
47	A	90.6	85
48	A	89.8	79
49	B	83.6	76
50	D	59.7	63
51	A	93.4	87
52	A	65.9	54
53	C	43.7	42
54	A	72.0	65
55	C	82.8	78
56	B	83.0	77
57	C	62.3	55
58	D	78.7	76
59	A	57.1	52
60	D	57.9	59
61	A	48.2	46
62	C	65.0	60
63	A	54.5	44
64	B	62.9	59
65	C	43.9	42
66	D	76.0	72
67	A	50.9	44
68	D	53.9	41
69	D	27.6	17
70	D	38.6	30
71	B	34.6	37
72	A	45.0	33
73	B	61.5	55
74	D	72.7	63
75	C	45.5	33

The instrument was developed by researchers at Harvard University Graduate School of Education to measure student learning in the so-called "intangible" areas of science. These intangibles include the nature of scientific inquiry, the role of scientists as inquirers, and their relationship to science as an institution in our technological society.

Table 5-3

GRADES 9 AND 12 STUDENT RESPONSES TO TOUS

	Grade 9			Grade 12		
	Boys	Girls	Total	Boys	Girls	Total
N	1234	1144	2426	1059	1034	2125
Average rawscore	21.9	22.6	21.8	24.6	25.8	25.3
Standard deviation	6.1	5.8	6.0	6.7	6.2	6.5

The data presented in Table 5-3 indicates that there is a growth of knowledge and understanding in this domain from grade nine to grade twelve. Whether the extent of the difference is enough cannot be answered within the framework of this study. However, the differences are significant ($p < 0.001$) given the numbers of students involved.

The TOUS test measures in the areas of understandings and, to some extent, opinions. The results are best examined on the basis of the objective Clusters #5, 6 and 7.

5.4 Grade Twelve Student Results by Science Background

Students were asked to report the number of 30-level science courses in which they were enrolled or for which they had received credit. Thus, the students reporting no science probably have had a 10- or 20-level course, but have had no driving interest in continuing in a science program.

Table 5-4

GRADE 12 STUDENT PERFORMANCE
BY SCIENCE COURSES: RAW SCORES

30-Level Science:	3 Sciences	2 Sciences	1 Science	No Science	Total Group
N	515	707	425	478	2125
$\overline{X}$ STEP	49.0	53.0	46.0	36.4	46.9
SD	14.6	9.0	8.7	9.4	18.3
$\overline{X}$ Supp.	6.9	7.6	6.5	5.6	6.7
SD	2.2	1.7	1.7	1.9	2.0
$\overline{X}$ TOUS	25.6	27.1	25.6	21.9	25.3
SD	7.0	6.1	5.8	5.7	6.5
$\overline{X}$ TOTAL	81.6	87.7	78.1	63.8	78.9
SD	21.5	13.7	13.0	14.2	18.2

The students enrolled in two 30-level sciences out-performed students in any other enrollment pattern. Those students taking three 30-level courses did not show up as well as their peers in the two 30-level pattern. Those students not taking any 30-level science did significantly less well on all tests. Their poorest performance was on the standardized STEP test where they trailed their colleagues in all other enrollment patterns.

Table 5-5

TEST ON DIFFERENCES BETWEEN MEANS
USING STUDENTS' t (GRADE 12)

*** 0.001 LEVEL OF CONFIDENCE
** 0.01 LEVEL OF CONFIDENCE

		2 Sciences	1 Science	No Science	Total Group
3 Sciences	STEP	***	***	***	**
	SUPP.	***	***	***	
	TOUS	***		***	
	TOTAL	***		***	**
2 Sciences	STEP		***	***	***
	SUPP.		***	***	***
	TOUS		***	***	***
	TOTAL		***	***	***
1 Science	STEP			***	
	SUPP.			***	**
	TOUS			***	
	TOTAL			***	
No Science	STEP				***
	SUPP.				***
	TOUS				***
	TOTAL				***

The test for significance is a ratio of the difference between two scores and the standard error of measurement. In this case, the table shows that the "no science" students did much more poorly than any other group and that the "2-science" group did much better than any other group.

5.5 Grade Twelve Results by Sex of Student

Table 5-6

GRADE 12 STUDENT PERFORMANCE
BY SEX: RAW SCORES

	Male	Female	Mean Difference	"Student's t"	Confidence Level
N	1,059	1,031			
$\overline{X}$ STEP	48.90	44.88	4.02	7.54	.001+
SD	12.60	11.80			
$\overline{X}$ TOUS	24.62	25.78	0.92	3.25	.01+
SD	6.71	6.18			
$\overline{X}$ TOTAL	80.62	77.28	3.34	4.21	.001+
SD	18.90	17.43			

As can be seen from the table, boys did significantly better on the STEP test, girls did slightly better on the TOUS and, overall, boys did far better than girls.

5.6 Size of School as a Variable

The sample schools were classified by their reported grade twelve enrollment. Table 5-8 shows the average student performance on the test for each classification. Although none of the differences are significant, the pattern of differences in the means and standard deviations among the various sizes of schools may well indicate an area for further study.

Table 5-7

GRADE 12 STUDENT PERFORMANCE
BY SCHOOL SIZE: RAW SCORES

	>30		31-60		61-100		>100		Total	
N	10		5		4		5		24	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
STEP	46.7	11.5	46.1	11.7	46.3	11.9	45.5	11.6	46.9	18.3
TOUS	26.6	5.7	25.2	5.7	25.7	5.5	24.0	6.7	25.3	6.5
TOTAL	80.3	17.1	77.9	16.9	78.8	16.7	76.2	18.7	78.9	18.2

Differences are not significant ($p > 0.05$).

Table 5-8

GRADE 12 RESULTS BY ITEM CLUSTER

Item Cluster	Average (%)	Items	Rank
1. Biology	72.7	16	1
2. Physics	59.7	15	5
3. Chemistry	51.8	17	7
4. Earth Science	47.9	3	8
5. Methods of Science	64.8	36	3
6. Scientific Enterprise	60.1	13	4
7. Scientists	67.6	13	2
8. Sciencing	52.1	19	6

5.7 Grade Twelve Objective-Based Results

The referenced criteria were the general objectives or goals of secondary science taken from the Alberta Program of Studies for Senior High School--1975, p. 186-187. Each item on the three-part test has been matched with a cluster or group of related program objectives. The eight clusters are:

1. Biology concepts: Cell biology, classification, ecology, genetics and biological processes.

2. Physics concepts: Motion, vectors, conservation laws, waves, fields and energy.

3. Chemistry concepts: Properties of matter, atomic theory, organic reactions, redox and proton chemistry.

4. Earth-Science concepts: Crust, atmosphere, solar system.

5. Scientific methods and thought processes: Inferring, hypothesizing, interpreting data and experimental procedures.

6. Scientific enterprise: Science as a human activity, role of instrumentation and resources, international character of science.

7. Scientists: Generalizations about scientists as people and as reseachers.

8. Sciencing: General concepts about the procedures used in science, the aims and limitations of science, the interrelationship of science and technology.

For each cluster the number of items have been identified, the average student response to the item has been reported, and the average response for the cluster (with its standard deviation) has been reported. Categorization was done mainly on the basis of the technical reports from the test publishers.

Sample items have also been included with each cluster to clarify for the reader how the data were collected.

5.7.1 Item Cluster #1 - Biology Concepts

For the convenience of the reader the content dimension has been clustered according to the course content.

Table 5-9

AVERAGE GRADE 12 STUDENT PERFORMANCE ON ITEM CLUSTER #1

Content Area Of Items		# Of Items	Student Average (%)
GRADE 10	- Cell biology - Classification of organisms	5	78.8
GRADE 11	- Ecology - Genetics	5	72.1
GRADE 12	- Biological Processes	6	68.1
		16	

Item cluster response - 11.6/16
Average student response - 72.7%
Item cluster standard deviation - 15.8%

The average student response was 11.63 out of 16 items or 72.7 percent mastery of the items in this cluster. This performance level ranks this cluster as first out of eight. One could suggest that this showing results from a reinforcing of school learning by observation of the environment. For example, biology, having been taught from the content around the student, has created an environmental awareness that is continually reinforced.

Sample Items - #1, 11, 79

1. The usual biological classification of plants and animals is based primarily on

a. age.
b. size.
c. geographical distribution.
• d. structure.

11. Chlorophyll-containing organisms were recently discovered almost 3 miles below the surface of the Mediterranean. Which of the following makes this discovery unusual?

a. The pressure at this depth
b. The high temperature at this level
• c. The near absence of light at this depth
d. The lowered concentration of carbon dioxide in the ocean as compared with the atmosphere

79. In his paper, "On the antibacterial action of cultures of a Penicillium", written in 1929, Alexander Fleming reported on the curious effect that a culture of mould spores had on staphylococcus colonies. As a result of this laboratory accident,

a. Alexander Fleming was fired.
b. more careful laboratory procedures were instituted.
c. more people became infected with staphylococcus.
• d. a life-saving drug was discovered.

5.7.2 Item Cluster #2 - Physics Concepts

The concepts were grouped according to their appearance in the newly revised course. The students would have dealt with them in a different order.

Table 5-10

AVERAGE GRADE 12 STUDENT PERFORMANCE
ON ITEM CLUSTER #2

Content Area Of Items		# Of Items	Student Average (%)
GRADE 10	- Motion - Acceleration - Newtonian mechanics, force, mass, inertia - Vectors	6	71.4
GRADE 11	- Conservation laws, mass and energy - Conservation of momentum - Waves - Energy, work, power and heat	3	54.3
GRADE 12	- Light: Nature and geometry - Electromagnetic radiation - Fields- electrical, magnetic and electromagnetic - Structure of matter	6	51.8
		15	

Item cluster response - 8.9/15
Average student response - 59.7%
Item cluster standard deviation - 16.3%

As one may have expected, many high school students performed rather poorly on physics knowledge and understandings. This cluster ranked fifth of the eight.

Sample Items - #4, 24

4. A motorist near the equator drives 3 miles east, then 4 miles south, and finally 3 miles west. At the end of his trip he is

a. 10 miles south of his starting point.
b. 10 miles southeast of his starting point.
• c. 4 miles south of his starting point.
d. 4 miles southeast of his starting point.

24. In a particular power plant, fuel is burned to boil water. The steam is used to drive a turbine which in turn drives a generator. Which of the following energy transfers is NOT involved?

a. Mechanical to electrical
• b. Electrical to chemical
c. Chemical to thermal
d. Thermal to mechanical

5.7.3 Item Cluster #3 - Chemistry Concepts

These concepts have been arranged in the order in which they appear in the current chemistry program--not necessarily in the order that the students dealt with them.

Table 5-11

AVERAGE GRADE 12 STUDENT PERFORMANCE ON ITEM CLUSTER #3

Content Area Of Items		# Of Items	Student Average (%)
GRADE 10	- Properties of matter - Atomic theory, symbols, formulae and equations	8	52.8
GRADE 11	- Solutions - Bonding - Organic chemistry	6	62.2
GRADE 12	- Chemical change and energetics - Proton chemistry - Redox reactions	3	44.9
		17	

Item cluster response - 8.8/17
Average student response - 51.8%
Item cluster standard deviation - 10.9%

Chemistry, along with physics, showed up as being less well understood by the general high school student population. The average student response ranked this cluster as seventh out of eight. This is a surprisingly low ranking for what is the second most popular science subject.

Sample Items - #6, 23, 26

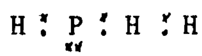
6. The use of mercury in laboratory thermometers depends on which of the following?

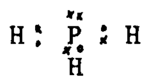
I. Its density is 13.6 grams/ml.
II. Its melting point is -39 1/4 C.
III. Its change in volume with change in temperature is relatively uniform.

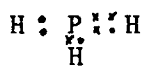
a. I only
b. II only
c. I and II only
• d. II and III only

23. Since a phosphorus atom has 5 electrons in its outer energy level, a molecule of PH_3 would be expected to have which of the following structures?

a. 

b. 

• c. 

d. 

26. The formation of chemical bonds is usually described in terms of which of the following particles?

a. Protons
• b. Electrons
c. Neutrons
d. Alpha particles

5.7.4 Item Cluster #4 - Earth Science Concepts

These concepts are not taught in the senior high school program but there should be some retention of these selected concepts.

Table 5-12

AVERAGE GRADE 12 STUDENT PERFORMANCE
ON ITEM CLUSTER #4

Content Area Of Items	# Of Items	Student Average (%)
Knowledge about the earth's crust	2	38.3
Knowledge about the earth's atmosphere	1	67.0
	3	

Item cluster response - 1.4/3
Average student response to all items - 47.9%
Item cluster standard deviation - 17.4%

Since the concepts related to weather and climate are reinforced in the high school social studies program, it is not surprising that the observed pattern has emerged. The study of rocks is a very personal concern and is left to the individual to pursue. This cluster ranked at the bottom of the list of eight item clusters.

Sample Item - #41

41. Caves are most likely to be found in rock that

a. is very dense.
• b. is soluble in ground water.
c. has been changed by high pressure.
d. has been changed by high temperature.

5.7.5 Item Cluster #5 - Scientific Methods and Thought Processes

This is an important area of the curriculum that has received a great deal of emphasis in the past few years, particularly at the junior high school level.

Table 5-13

AVERAGE GRADE 12 STUDENT PERFORMANCE
ON ITEM CLUSTER #5

Content Area Of Items	# Of Items	Student Average (%)
Scientific Processes:		
Inferring, hypothesizing and predicting	9	64.7
Interpreting data	18	59.5
Experimental procedure	7	70.6
Classifying and observing	2	91.6
	36	

Item cluster response - 23.2/36
Average student response - 64.8%
Item cluster standard deviation - 83.3%

The student performance ranked this cluster third out of the eight objectives. That is, items calling for the knowledge and use of the scientific processes have been answered well enough to rank this cluster fairly highly. It is encouraging to note the relative ranking of the item cluster in light of the emphasis this area has received in the secondary science program.

Sample Items - #56, 62, 66

56. To gather data on the speed of a toy car, one would find which of the following the most accurate?

a. A balance, a set of weights, and a metrestick
• b. A tape measure and a stop watch
c. A yardstick and an oven timer
d. An odometer (mileage indicator from the dashboard of a car) and an alarm clock

62. Which of the following observations would support the conclusion that the solid is a mixture?

a. The solid can be ground to a fine powder.
b. The density of the solid is 2.25 grams/ml.
• c. The solid begins to melt at 300 1/4 C but is not completely melted until 375 1/4 C.
d. The solid is white in color.

66. Which of the following is NOT an advantage of using aspirin for medical purposes?

a. It is inexpensive.
b. It relieves pain rapidly.
c. It reduces inflammation in joints.
• d. It induces allergic hypersensitivity in many people.

5.7.6 Item Cluster #6 - Understanding About the Scientific Enterprise

This cluster has received more attention in the newly adopted science programs in chemistry and physics.

Table 5-14

AVERAGE GRADE 12 STUDENT PERFORMANCE ON ITEM CLUSTER #6

Content Area Of Items	# Of Items	Student Average (%)
Communication among scientists	3	67.4
Instrumentation	1	57.4
Money and resources	1	54.2
International character	1	84.0
Interaction of science and society	2	41.4
	8	

Item cluster response - 4.8/8
Average student response - 60.1%
Item cluster standard deviation - 19.3%

Not too surprisingly, this cluster falls in the middle rank in student achievement. The curriculum often ignores the objectives addressed to the larger view of science in society in the interests of learning more "science".

Sample Items - #94, 123

94. The scientists of today can work on more complex problems than the scientists of the past mainly because they

a. work much harder than earlier scientists.
b. have more imagination than earlier scientists.
• c. build on the work of earlier scientists.
d. are more intelligent than earlier scientists.

123. A young person who wants to have a career in science would be able to find a job as a scientist

a. only in the United States or Russia.
b. only in countries in American and Europe.
• c. in most countries of the world.
d. only in countries with large industries.

5.7.7 Item Cluster #7 - Understanding About Scientists

Table 5-15

AVERAGE GRADE 12 STUDENT PERFORMANCE
ON ITEM CLUSTER #7

Content Area Of Items	# Of Items	Student Average (%)
Generalizations about scientists as people	5	72.2
Institutional pressure on scientists	1	90.9
Abilities needed by scientists	5	65.3
	11	

Item cluster response - 7.4/11
Average student response - 67.6%
Item cluster standard deviation - 16.5%

The student response pattern shows a better-than-average understanding of scientists. This cluster ranked second of the eight objective clusters. It would seem that there is an awareness of scientists in our society and an appreciation of the role that they play in our technological society.

Sample Items - #104, 119

104. Mary writes down all the details of her science experiments. If Mary becomes a scientist, this training will help her to

a. be patient in doing her experiments.
• b. make better reports about her experiments.
c. think up theories from her experiments.
d. work out new experiments to perform.

119. In trying to decide whether or not to go to a new movie, a scientist would probably ask himself:

a. "Does the film use experimental techniques?"
b. "Is there cause and effect in the plot?"
• c. "Will I like this film?"
d. "Is the plot factual and accurate?"

5.7.8 Item Cluster #8 - Understandings About Science and Sciencing

This cluster consists of items that ask for responses about perceptions in the realm of goals and limitations on investigations in science.

Table 5-16

AVERAGE GRADE 12 STUDENT PERFORMANCE
ON ITEM CLUSTER #8

Content Area Of Items	# Of Items	Student Average (%)
Scientific methods:		
Tactics and strategies	9	41.1
Theories and models	7	63.4
Aims of science	6	56.8
Science and technology	2	39.7
Unity and interdependence of the sciences	2	60.1
	26	

Item cluster response - 13.5/26
Average student response - 52.1%
Item cluster standard deviation - 19.8%

Knowledge of what scientists do appears to be well in hand. But there seem to be some misconceptions about the aims of the scientific community and the role of science in pushing back the frontiers of knowledge.

Sample Items - #103, 120, 122

103. The chief purpose of the science of botany is to

a. find out what plants grow best in various kinds of soil.
• b. understand how plants live, grow, and reproduce.
c. develop new drugs and medicines from plants.
d. help farmers grow bigger plants and produce more food.

120. When we say that a scientist forms a <u>hypothesis</u> about an experiment, we mean that he

• a. makes a careful guess about what will happen.
b. gives directions for doing the experiment properly.
c. suggests how to make exact measurements.
d. describes how the experiment was carried out.

122. A scientific law describes

a. rules which scientists must obey.
• b. relationships between events in nature.
c. directions for doing good experiments.
d. good guesses about how things happen.

Date Due

Treasure, Morris - Alberta science achievement study: condensed report